

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

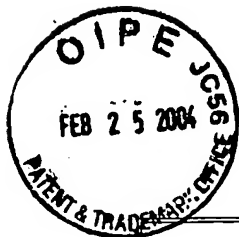
Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



PATENT  
ATTORNEY DOCKET NO. 50287/007002

Certificate of Mailing: Date of Deposit February 23, 2004

I hereby certify under 37 C.F.R. § 1.8(a) that this correspondence is being deposited with the United States Postal Service as **first class mail** with sufficient postage on the date indicated above and is addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Rosemarie Perullo

Printed name of person mailing correspondence

Signature of person mailing correspondence

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Sakari Kauppinen et al.	Art Unit:	Not yet assigned
Serial No.:	10/690,487	Examiner:	Not yet assigned
Filed:	October 21, 2003	Customer No.:	21559
Title:	Oligonucleotides Useful for Detecting and Analyzing Nucleic Acids of Interest		

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

SUBMISSION OF PRIORITY DOCUMENT

In connection with the above-referenced application, Applicants submit herewith a certified copy of Danish Patent Application PA 2003-00752, as filed on May 19, 2003, with the Danish Patent Office. Applicants petition for any necessary extensions of time for submission of this document. In addition, if there are any charges or any credits, please apply them to Deposit Account No. 03-2095.

Respectfully submitted,

Date:

February 23, 2004

Kristina Bieker-Brady, Ph.D., P.C.  
Reg. No. 39,109

Clark & Elbing LLP  
101 Federal Street  
Boston, MA 02110  
Telephone: 617-428-0200  
Facsimile: 617-428-7045





# Kongeriget Danmark

Patent application No.: PA 2003 00752

Date of filing: 19 May 2003

Applicant: Exiqon A/S  
(Name and address) Bygstubben 9  
DK-2950 Vedbæk  
Denmark

Title: Oligonucleotides useful for detecting and analyzing nucleic acids of interest

IPC: -

This is to certify that the attached documents are exact copies of the above mentioned patent application as originally filed.



**Patent- og Varemærkestyrelsen**  
Økonomi- og Erhvervsministeriet

02 October 2003

Pia Høybye-Olsen



**PATENT- OG VAREMÆRKESTYRELSEN**

19 MAJ 2003

Modtaget

1

# Oligonucleotides Useful for Detecting and Analyzing Nucleic Acids of Interest

The invention relates to nucleic acids and methods for expression profiling of mRNAs, identifying and profiling of particular mRNA splice variants, and detecting mutations, deletions, or duplications of particular exons, e.g., alterations associated with a disease such as cancer, in a nucleic acid sample, e.g., a patient sample.

## Background of the Invention

The field of the invention is oligonucleotides (e.g., oligonucleotide arrays) that are useful for detecting nucleic acids of interest and for detecting differences between nucleic acid samples (e.g., such as samples from a cancer patient and a healthy patient).

DNA chip technology utilizes miniaturized arrays of DNA molecules immobilized on solid surfaces for biochemical analyses. The power of DNA microarrays as experimental tools relies on the specific molecular recognition via complementary base-pairing, which makes them highly useful for massive parallel analyses. In the post-genomic era, microarray technology has thus become the method of choice for many hybridization-based assays, such as expression profiling, SNP detection, DNA re-sequencing, and genotyping on a genomic scale.

Expression microarrays are capable of profiling gene expression patterns of tens of thousands of genes in a single experiment. Hence, this technology provides a powerful tool for deciphering complex biological systems, and thereby greatly facilitates research in basic biology and living processes, as well as disease diagnostics, theranostics, and drug development. In a typical cell, the mRNAs are distributed in three frequency classes: (i) superprevalent (10-20% of the total mRNA mass); (ii) intermediate (40-45%); and (iii) low-abundant mRNAs (40-45%). It is therefore of utmost importance that the dynamic range and sensitivity of the expression arrays are optimal, especially when analyzing expression levels of messages or mRNA splice variants belonging to the low-abundant class.

The recent explosion of interest in DNA microarray technology has been sparked by two key innovations. The first was the use of non-porous solid support, such as glass or polymer as opposed to nylon or nitrocellulose filters, which has facilitated miniaturization and fluorescence-based detection. Roughly 20,000 cDNAs can be robotically spotted onto a microscope slide and hybridized with a double-labeled probe. The second was the development of methods for high-density spatial synthesis of oligonucleotides. The two key array technologies are outlined in the following.

## *Oligonucleotide arrays*

An efficient strategy for oligonucleotide microarray manufacturing involves DNA synthesis on solid surfaces using combinatorial chemistry. Most of the current technology is

developed by Affymetrix and Rosetta Inpharmatics. Glass is currently preferred as the synthesis support because of its inert chemical properties and low level of intrinsic fluorescence as well as the ability to chemically derivatize the surface. Of the three approaches currently used to manufacture oligonucleotide arrays, the light-directed deprotection method is the most effective one in generating high density microchips. A single round of synthesis involves light-directed deprotection, followed by nucleotide coupling. Photolithographic masking is used to control the regions of the chip designated for illumination. Affymetrix uses a combination of photolithography and combinatorial chemistry to manufacture its GeneChip Arrays. Using technologies adapted from the semiconductor industry, GeneChip manufacturing begins with a 5-inch square quartz wafer. Initially the quartz is washed to ensure uniform hydroxylation across its surface. The wafer is placed in a bath of silane, which reacts with the hydroxyl groups of the quartz and forms a matrix of covalently linked molecules. The distance between these silane molecules determines the probes' packing density, allowing arrays to hold over 500,000 features within 1.28 square centimeters. The principal disadvantage of this method is that a significant amount of chip design work and cost is associated with the mask design. Once a set of masks has been made, a large number of chips can be produced at a reasonable cost. The current pricing of oligonucleotide arrays available from Affymetrix are in the range of 5-10 fold more expensive than cDNA microarrays.

DNA-DNA hybridization using oligonucleotide chips is clearly different from that of cDNA microarrays. Hybridizations involving oligos are much more sensitive to the GC content of individual heteroduplexes. In addition, single base mismatches have a pronounced effect on the hybridization reassociation of short oligos, and point mutations can thus be readily detected using oligo chips.

#### *cDNA microarrays*

cDNA microarrays containing large DNA segments such as cDNAs are generated by physically depositing small amounts of each DNA of interest onto known locations on glass surfaces. Two technologies for printing microarrays are (1) mechanical microspotting, and (2) ink-jetting. Mechanical microspotting has been extensively used at, e.g., Stanford University, and it utilizes pins or capillaries to deposit small quantities of DNA onto known addresses using motion control systems. Recent advances in microspotting technology using modern arraying robots allow for the preparation of 100 microarrays containing over 10,000 features in less than 12 hours. A DNA arrayer is relatively easy to set up, and the cost is usually low compared to on-chip oligoarrayers. cDNA microarrays are capable of profiling gene expression patterns of tens of thousands of genes in a single experiment. To compare the relative abundance of the arrayed gene sequences in two DNA or RNA samples, e.g., the total mRNA isolated from two

different cell populations, the two samples are first labeled using two different fluorescent dyes such as Cy-3 and Cy-5. The labeled samples are mixed and hybridized to the clones on the array slide. After the hybridization, laser excitation of the incorporated, fluorescent target molecules yields an emission with a characteristic spectra, which is measured with a confocal laser scanner. The monochrome images from the scanner are imported to the software in which the images are pseudo-colored and merged. Data from a single hybridization is viewed as a normalized ratio in which significant deviations from the ratio are indicative of either increased or decreased expression levels relative to the reference sample. Data from multiple experiments can be examined using any number of data mining tools.

#### 10 *Current status of array technology*

It has now become clear that cDNA microarrays, originally developed by Pat Brown and co-workers at the Stanford University, are sensitive, but may not be sufficiently specific with respect to, e.g., discrimination of homologous transcripts in gene families and alternatively spliced isoforms. On the other hand, the Affymetrix GeneChip system is specific, but may not be sensitive enough. This lack of sensitivity may explain why Affymetrix uses 16x26-mer perfect match capture probes together with 16x25-mer mismatch probes per transcript in its expression profiling chips resulting in enormous data sets in genome-wide arrays. Therefore, the functional genomics field is in the process of switching, as they run out of samples, from existing PCR-amplified cDNA fragment libraries for microarraying to custom longer oligonucleotide arrays comprising transcript-specific oligonucleotide capture probes typically in the range of 30-mers to 80-mers, thus addressing both specificity and sensitivity.

#### *Alternative splicing*

As the field of genomics research is shifting from the acquisition of genome sequences to high-throughput functional genomics, there is an increasing need to understand the dynamics within the genetic regulation as well as RNA and protein sequences in order to elucidate gene expression in all its complexity. A common feature for eukaryotic genes is that they are composed of protein-encoding exons and introns. Introns (intra-genic-regions) are non-coding DNA which interrupt the exons. Introns are characterized by being excised from the pre-mRNA molecule in RNA splicing, as the sequences on each side of the intron are spliced together. RNA splicing not only provides functional mRNA, but is also responsible for generating additional diversity. This phenomenon is called alternative splicing, which results in the production of different mRNAs from the same gene. The mRNAs that represent isoforms arising from a single gene can differ by the use of alternative exons or retention of an intron that disrupts two exons. This process often leads to different protein products that may have related or drastically different, even antagonistic, cellular functions. There is increasing evidence

indicating that alternative splicing is very widespread (Croft *et al.* Nature Genetics, 2000). Recent studies have revealed that at least 60% of the roughly 35,000 genes in the human genome are alternatively spliced. Clearly, by combining different types of modifications and thus generating different possible combinations of transcripts of different genes, alternative splicing is a potent mechanism for generating protein diversity. Analysis of the spliceome, in turn, represents a novel approach to both functional genomics and pharmacogenomics.

#### *Antisense transcription in eukaryotes*

RNA-mediated gene regulation is widespread in higher eukaryotes and complex genetic phenomena like RNA interference, co-suppression, transgene silencing, imprinting, methylation, and possibly position-effect variegation and transvection, all involve intersecting pathways based on or connected to RNA signalling (Mattick 2001; EMBO reports 2, 11: 986-991). Recent studies indicate that antisense transcription is a very common phenomenon in the mouse and human genomes (Okazaki *et al.* 2002; Nature 420: 563-573; Yelin *et al.* 2003, Nature Biotechnol.). Thus, antisense modulation of gene expression in e.g. human cells may be a common regulatory mechanism. In light of this, the present invention provides novel tools, in which non-naturally occurring nucleic acids, such as LNA oligonucleotides, can be designed to silence or modulate the regulation of a given mRNA by non-coding antisense RNA, by designing a complementary sense LNA oligonucleotide for the regulatory antisense RNA. This has a high potential in target identification, target validation and therapeutic use of LNA oligonucleotides as modulating and silencing sense nucleic acid agents.

#### *Misplaced control of alternative splicing can cause disease*

The detection of the detailed structure of all transcripts is an important goal for molecular characterization of a cell or tissue. Without the ability to detect and quantify the splice variants present in one tissue, the transcript content or the protein content cannot be described accurately. Molecular medical research shows that many cancers result in altered levels of splice variants, so an accurate method to detect and quantify these transcripts is required. Mutations that produce an aberrant splice form can also be the primary cause of such severe diseases such as spinal muscular dystrophy and cystic fibrosis.

Much of the study of human disease, indeed much of genetics is based upon the study of a few model organisms. The evolutionary stability of alternative splicing patterns and the degree to which splicing changes according to mutations and environmental and cellular conditions influence the relevance of these model systems. At present, there is little understanding of the rates at which alternative splicing patterns change, and the factors influencing these rates. Table I shows a set of genes that are known to be alternatively spliced and that are orthologs of known human disease genes.

Table 1. *C. elegans* disease orthologs that are known to be differentially spliced in *C. elegans*.

Disease	<i>C. elegans</i> gene	BLAST E value
brABL1	M79.1A	1.00E-162
X-Linked Lymphoprol.-SH2D1A	M79.1A	2.00E-58
Cyclin Dep. Kinase 4-CDK4	F18H3.5A	1.00E-124
HNPCC*-PMS2	H12C20.2A	1.00E-123
Neurofibromatosis 2-NF2	C01G8.5A	5.00E-163
Duchenne MD+-DMD	F32B4.3A	0.00E+00
Coffin-Lowry-RPS6KA3	T01H8.1A	2.00E-13
Septooptic Dysplasia-HESX1	Y113G7A.6A	1.00E-152
Non-Insulin Dep. Diabet.-PCSK1	F11A6.1A	1.00E-166
Bartter's-SLC12A1	Y37A1C.1A	1.00E-167
Gitelmans-SLC12A3	Y37A1C.1A	0.00E+00
Hered. Spherocytosis-ANK1	B0350.2A	1.00E-09
Darier-White-SERCA	K11D9.2A	0.00E+00
Spondyloepip.Dysp.-COL2A1	F01G12.5A/let-2	9.00E-20

Previously, other microarray analyses have been performed with the aim of detecting either  
5 splicing of RNA transcripts *per se* in yeast, or of detecting putative exon skipping splicing  
events in rat tissues, but neither of these approaches had sufficient resolution to estimate  
quantities of splice variants, a factor that could be essential to an understanding of the changes  
in cell life cycle and disease.

Thus, improved methods are needed for nucleic acid amplification, hybridization, and  
10 classification. Desirable methods can distinguish between mRNA splice variants and quantitate  
the amount of each variant in a sample. Other desirable methods can detect differences in  
expressions patterns between patient nucleic acid samples and nucleic acid standards.

#### Summary of the Invention

15 The present invention demonstrates the usefulness of LNA-modified oligonucleotides in  
the construction of highly specific and sensitive microarrays for expression profiling (e.g.,  
mRNA splice variant detection) and comparative genomic hybridization. The invention  
provides novel technology platforms based on nucleic acids with LNA or other high affinity

nucleotides for sensitive and specific assessment of alternative splicing using microarray technology. As opposed to high-density cDNA or DNA oligonucleotide microarrays, LNA microarrays are able to discriminate between highly homologous as well as differentially spliced transcripts. The present methods greatly facilitate the analysis of gene expression patterns from a particular species, tissue, cell type. The analysis of the human spliceome provides important information for pharmacogenetics. Thus, the present methods are highly valuable in medical research and diagnostics as well as in drug development and toxicological studies.

In general, the invention features populations of high affinity nucleic acids that have duplex stabilizing properties and thus are useful for a variety of nucleic acid detection, amplification, and hybridization methods (e.g., expression or mRNA splice variant profiling). Some of these oligonucleotides contain novel nucleotides created by combining specialized synthetic nucleobases with an LNA backbone, thus creating high affinity oligonucleotides with specialized properties such as reduced sequence discrimination for the complementary strand or reduced ability to form intramolecular double stranded structures. The invention also provides improved methods for identifying nucleic acids in a sample and for classifying a nucleic acid sample by comparing its pattern of hybridization to an array to the corresponding pattern of hybridization of one or more standards to the array (e.g., comparative genomic hybridization).

Other desirable modified bases have decreased ability to self-anneal or to form duplexes with oligonucleotides containing one or more modified bases. The invention also provides arrays of nucleic acids containing these modified bases that have a decreased variance in melting temperature and/or an increased capture efficiency compared to naturally-occurring nucleic acids. These arrays as well as the oligonucleotides in solution can be used in a variety of applications for the detection, characterization, identification, and/or amplification of one or more target nucleic acids. These oligonucleotides and oligonucleotides of the invention in general can also be used for solution assays, such as homogeneous assays.

#### *Merged Probes*

In one aspect, the invention features a non-naturally-occurring nucleic acid with a melting temperature that is at least 3, 5, 8, 10, 12, 15, 20, 25, 30, 35, or 40°C higher than that of the corresponding control nucleic acid with 2'-deoxynucleotides. The nucleic acid hybridizes to a first region within a first exon of a target nucleic acid and to a second region within a second exon of the target nucleic acid that is adjacent to the first exon.

In a related aspect, the invention provides a non-naturally-occurring nucleic acid with a melting temperature that is at least 3, 5, 8, 10, 12, 15, 20, 25, 30, 35, or 40°C

higher than that of the corresponding control nucleic acid with 2'-deoxynucleotides. The nucleic acid hybridizes to a first region within an exon of a target nucleic acid and to a second region within an intron of the target nucleic acid that is adjacent to the exon.

In another aspect, the invention features a non-naturally-occurring nucleic acid 5 with a melting temperature that is at least 3, 5, 8, 10, 12, 15, 20, 25, 30, 35, or 40°C higher than that of the corresponding control nucleic acid with 2'-deoxynucleotides. The nucleic acid hybridizes to a first region within a first intron of a target nucleic acid and to a second region within a second intron of the target nucleic acid that is adjacent to the first intron.

10 In yet another aspect, the invention provides a nucleic acid that is a non-naturally-occurring nucleic acid with a capture efficiency that is at least 10, 25, 50, 100, 150, 200, 500, 800, 1000, or 12000% greater than that of a corresponding control nucleic acid with 2'-deoxynucleotides at the temperature equal to the melting temperature of the nucleic acid. The nucleic acid hybridizes to a first region within a 15 first exon of a target nucleic acid and to a second region within a second exon of the target nucleic acid that is adjacent to the first exon.

In a related aspect, the invention features a nucleic acid that is a non-naturally-occurring nucleic acid with a capture efficiency that is at least 10, 25, 50, 100, 150, 200, 500, 800, 1000, or 12000% greater than that of a corresponding control nucleic 20 acid with 2'-deoxynucleotides at the temperature equal to the melting temperature of the nucleic acid. The nucleic acid hybridizes to a first region within an exon of a target nucleic acid and to a second region within an intron of the target nucleic acid that is adjacent to the exon.

In yet another aspect, the invention provides a nucleic acid that is a non- 25 naturally-occurring nucleic acid with a capture efficiency that is at least 10, 25, 50, 100, 150, 200, 500, 800, 1000, or 12000% greater than that of a corresponding control nucleic acid with 2'-deoxynucleotides at the temperature equal to the melting temperature of the nucleic acid. The nucleic acid hybridizes to a first region within a first intron of a target nucleic acid and to a second region within a second intron of the 30 target nucleic acid that is adjacent to the first intron.

In desirable embodiments, the nucleic acids of the invention featuring a non-naturally occurring nucleic acid exhibit increased duplex stability due to slower rates of dissociation of the nucleic acid complexes (the off-rate) (Christensen *et al.* 2001, *Biochem. J.* 354: 481-484).



In still another aspect, the invention features a nucleic acid that is an LNA (i.e., a nucleic acids with one or more LNA units) and that hybridizes to a first region within a first exon of a target nucleic acid and to a second region within a second exon of the target nucleic acid that is adjacent to the first exon.

5 In another aspect, the invention features a nucleic acid that is an LNA and that hybridizes to a first region within an exon of a target nucleic acid and to a second region within an intron of the target nucleic acid that is adjacent to the exon.

In one aspect, the invention provides nucleic acid that is an LNA and that hybridizes to a first region within a first intron of a target nucleic acid and to a second  
10 region within a second intron of the target nucleic acid that is adjacent to the first intron.

In desirable embodiments of any of the above aspects, the length of the segment of the nucleid acid hybridizing to the first region and the length of the segment of the nucleic acid hybridizing to the second region are between 3 and 50 nucleotides, 10 and 40 nucleotides, or 20 and 30 nucleotides, inclusive. The length of the segment of the  
15 nucleid acid hybridizing to the first region and the length of the segment of the nucleid acid hybridizing to the second region may be the same length or different lengths. Desirably, the nucleic acid contains LNA units are are symmetrically spaced on both sides of a junction between either two exons, an exon and an intron, or two introns. Desirably, the nucleic acid has one or more LNA units within 5, 4, 3, 2, or 1 nucleotides  
20 of a junction between either two exons, an exon and an intron, or two introns.

In another aspect, the invention features a population of nucleic acids that includes one or more nucleic acids of any one of the above aspects.

#### *Internal Probes*

In another aspect, the invention features a non-naturally-occurring nucleic acid  
25 with a melting temperature that is at least 3, 5, 8, 10, 12, 15, 20, 25, 30, 35, or 40°C higher than that of the corresponding control nucleic acid with 2'-deoxynucleotides. The nucleic acid hybridizes to only one exon or to only one intron of a target nucleic acid.

In a related aspect, the invention features a non-naturally-occurring nucleic acid  
30 with a capture efficiency that is at least 10, 25, 50, 100, 150, 200, 500, 800, 1000, or 12000% greater than that of a corresponding control nucleic acid with 2'-deoxynucleotides at the temperature equal to the melting temperature of the nucleic acid. The nucleic acid hybridizes to only one exon or to only one intron of a target nucleic acid.

In another aspect, the invention features a nucleic acid that is an LNA and that hybridizes to only one exon or to only one intron of a target nucleic acid.

In desirable embodiments of the above aspects for nucleic acids that hybridizes to only one exon or only one intron, the nucleic acid does not hybridize to both an exon and an intron.

In another aspect, the invention features a population of nucleic acids that includes one or more nucleic acids of any one of the above aspects.

#### *Pharmaceutical Compositions and Nucleic Acid Populations*

In another aspect, the invention features a pharmaceutical composition that includes one or more of the nucleic acids of the invention and a pharmaceutically acceptable carrier, such as one of the carriers described herein.

In another aspect, the invention features a population of two or more nucleic acids of the invention. The populations of nucleic acids of the invention may contain any number of unique molecules. For example, the population may contain as few as 10,  $10^2$ ,  $10^4$ , or  $10^5$  unique molecules or as many as  $10^7$ ,  $10^8$ ,  $10^9$  or more unique molecules. In desirable embodiments, at least 1, 5, 10, 50, 100 or more of the polynucleotide sequences are a non-naturally-occurring sequence. Desirably, at least 20, 40, or 60% of the unique polynucleotide sequences are non-naturally-occurring sequences. Desirably, the nucleic acids are all the same length; however, some of the molecules may differ in length.

#### *Desirable Embodiments of Any of the Above Aspects*

In desirable embodiments of any of the above aspects, the length of one or more nucleic acids (e.g., nucleic acids in a nucleic acid population of the invention) is between 15 and 150 nucleotides, 5 and 100 nucleotides, 20 and 80 nucleotides, or 30 and 60 nucleotides in length, inclusive. In particular embodiments, the nucleic acid is 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50 nucleotides or at least 60, 70, 80, 90, 100, 120, or 130 nucleotides in length. In additional embodiments, the nucleic acid is between 8 and 40 nucleotides, such as between 9 and 30, or 12 and 25, or 15 and 20 nucleotides. Desirably, at least 5, 10, 15, 20, 30, 40, 50, 60, or 70% of the nucleotides in the nucleic acid are LNA units. In desirable embodiments, every second nucleotide, every third, every fourth nucleotide, every fifth nucleotide, or every sixth nucleotide in the nucleic acid is an LNA unit. In various embodiments, (i) every second and every third nucleotide, (ii) every second and every fourth nucleotide, (iii) every second and every fifth nucleotide, (iv) every second and every sixth nucleotide, (v) every third and every fourth nucleotide, (vi) every third and every fifth nucleotide, (vii) every third and every sixth nucleotide, (viii) every fourth and every fifth nucleotide, (ix)

every fourth and every sixth nucleotide, and/or (x) every fifth and every sixth nucleotide in the nucleic acid is an LNA unit. Desirably, every second, every third, and every fourth nucleotide in the nucleic acid is an LNA unit. In desirable embodiments, the nucleic acids of the invention have one or more of the following substitution patterns 5 which is repeated throughout the nucleic acids: XxXx, xXxX, XxxXxx, xXxxXx, xxXxxX, XxxxXxxx, xXxxxXxx, xxXxxxXx, or xxxXxxxX in which "X" denotes an LNA unit and "x" denotes a DNA or RNA unit. In some embodiments, the nucleotides that are not LNA units are naturally-occurring DNA or RNA nucleotides.

In various embodiments, the nucleic acid comprises two or more contiguous 10 LNA units. Desirably, the nucleic acid comprises at least 2, 3, 4, 5, 6, 7, or 8 contiguous LNA units. In desirable embodiments, the number of contiguous LNA units is between 5 and 20% or 10 and 15% of the total length of the nucleic acid. In a particular embodiment, 5 contiguous nucleotides of a 50-mer merged probe are LNA units. In one embodiment, the nucleic acid does not have greatly extended stretches of modified 15 DNA or RNA residues, e.g. greater than about 4, 5, 6, 7, or 8 consecutive modified DNA or RNA residues. According to this embodiment, one or more non-modified DNA or RNA units are present after a consecutive stretch of about 3, 4, 5, 6, 7, or 8 modified nucleic acids.

Other desirable nucleic acids have an LNA substitution pattern that results in the 20 formation of negligible secondary structure by the nucleic acids with itself. In one such embodiment, the nucleic acids do not form hairpins or do not form other secondary structures that would otherwise inhibit or prevent their binding to a target nucleic acid. Desirably, opposing nucleotides in a palindrome pair or opposing nucleotides in inverted repeats or in reverse complements are not both LNA units. In various 25 embodiments, the nucleic acids in the first population form less than 3, 2, or 1 intramolecular base-pairs or base-pairs between two identical molecules. In desirable embodiments, the nucleic acid does not have LNA-5-nitroindole: LNA-5-nitroindole intramolecular base-pairs.

In other desirable embodiments, at least one LNA unit (e.g., at least 2, 3, 4, 5, 6, 30 7, 8, 9, or 10 LNA units) in the nucleic acid hybridizes to a first region within a first exon of a target nucleic acid and at least one LNA unit (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 LNA units) in the nucleic acid hybridizes to a second region within a second exon of the target nucleic acid that is adjacent to the first exon. The number of LNA units that bind to each region can be the same or different. In some embodiments, the 5' 35 terminal nucleotide of the nucleic acid is or is not an LNA unit. Desirably, the 3'

terminal nucleotide of the nucleic acid is not an LNA unit (e.g., the nucleic acid may contain a 3' terminal naturally-occurring nucleotide).

Desirably, the nucleic acid can distinguish between different nucleic acids (e.g., mRNA splice variants) that cannot be distinguished using a naturally-occurring control nucleic acid (e.g., a control nucleic acid that consists of only 2'-deoxynucleotides such as a control nucleic acid of the same length as the nucleic acid of the invention). Desirably, the hybridization intensity of the nucleic acid for an exon of interest is at least 2, 3, 4, 5, 6, or 10 fold greater than the hybridization intensity of the nucleic acid for another exon in the same target nucleic acid (e.g., mRNA) or in another nucleic acid. Desirably, the hybridization intensity of the nucleic acid for target nucleic acid is at least 2, 3, 4, 5, 6, or 10 fold greater than the hybridization intensity for a non-target nucleic acid with less than 99, 95, 90, 80, 70, or 60% sequence identity to the target nucleic acid.

Desirably, all of the nucleic acids of the population or all of the nucleic acids of a subpopulation of the population are the same length. In some embodiments, the population includes one or more nucleic acids of a different length. In some embodiments, longer nucleic acids contain one or more nucleotides with universal bases. For example, nucleotides with universal bases can be used to increase the thermal stability of nucleic acids that would otherwise have a thermal stability lower than some or all of the nucleic acids in the population. In some embodiments, one or more nucleic acids have a universal base located at the 5' or 3' terminus of the nucleic acid. In desirable embodiments, one or more (e.g., 2, 3, 4, 5, 6, or more) universal bases are located at the 5' and 3' termini of the nucleic acid. Desirably, all of the nucleic acids in the population have the same number of universal bases. Desirable universal bases include inosine, pyrene, 3-nitropyrrole, and 5-nitroindole.

In desirable embodiments, the nucleic acid has at least one LNA A or LNA T. In some embodiments, each nucleic acid has at least one LNA A or LNA T. Desirably, all of the adenine and thymine-containing nucleotides in the LNA are LNA A and LNA T, respectively. In some embodiments, a nucleic acid with an increased capture efficiency or melting temperature compared to a control nucleic acid has at least one LNA T or LNA C. In some embodiments, all of the thymidine and cytosine-containing nucleotides in the LNA are LNA T and LNA C, respectively. In some embodiments, a nucleic acid with an increased specificity or decreased self-complementarity compared to a control nucleic acid has at least one LNA A or LNA C. In some embodiments, all of the adenine and cytosine-containing nucleotides in the LNA are LNA A and LNA C,

respectively. Desirably, at least 10, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100% of the nucleic acids in the population have one or more LNA units.

In desirable embodiments, the LNA has at least one 2,6-diaminopurine, 2-aminopurine, 2-thio-thymine, 2-thio-uracil, inosine, or hypoxanthine base. Desirably, the LNA has a nucleotide with a 2'-O, 4'-C-methylene linkage between the 2' and 4' position of a sugar moiety. In some embodiments, one or more nucleic acids in the first population are LNA/DNA, LNA/RNA, or LNA/DNA/RNA chimeras.

In desirable embodiments of any of the above aspects, the variance in the melting temperature of the population is at least 10, 20, 30, 40, 50, 60, or 70% less than the variance in the melting temperature of the corresponding control population of nucleic acids of the same length with 2'-deoxynucleotides (e.g., DNA nucleotides) instead of LNA units or other modified or non-naturally-occurring units. In desirable embodiments, the standard deviation in melting temperature is less than 10, 9.5, 9, 8.5, 8, 7.5, 7, 6.5, or 6. In certain embodiments, the range in melting temperatures for nucleic acids in the population is less than 70, 60, 50, 40, 30, or 20°C. Desirably, the variance in the melting temperature of the population is less than 59, 50, 40, 30, 25, 20, 15, 10, or 5.

In still other embodiments, the nucleic acids are covalently bonded to a solid support. Desirably, the nucleic acids are in a predefined arrangement. In various embodiments, the first population has at least 10; 100; 1,000; 5,000; 10,000; 100,000; or 1,000,000 different nucleic acids. Desirably, the nucleic acids in the population together hybridize to at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, or 100% of the exons of a target nucleic acid. In desirable embodiments, the population includes nucleic acids that together hybridize to at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, or 100% of the nucleic acids expressed by a particular cell or tissue. In some embodiments, the population includes nucleic acids that together hybridize to at least one exon from at least 1, 5, 10, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100% of the nucleic acid sequences expressed by a particular cell or tissue at a given point in time (e.g., an expression array with sequences corresponding to the sequences of mRNA molecules expressed by a particular cell type or a cell under a particular set of conditions). In some embodiments, the plurality of nucleic acids are used as PCR primers or FISH probes.

Desirable modified bases of the present invention when incorporated into the central position of a 9-mer oligonucleotide (all other eight residues or units being natural DNA or RNA units with natural bases) exhibit a  $T_m$  difference equal to or less than about 15, 12, 10, 9, 8, 7, 6, 5, 4, 3 or 2°C upon hybridizing to the four complementary oligonucleotide variants that are

identical except for the unit corresponding to the LNA unit, where each variant has one of the natural bases uracil, cytosine, thymine, adenine or guanine. That is, the highest and the lowest  $T_m$  (referred to herein as the  $T_m$  differential) obtained with such four complementary sequences is 15, 12, 10, 9, 8, 7, 6, 5, 4, 3 or 2°C or less.

5 Modified nucleic acid oligomers of the invention desirably contain at least one LNA unit, such as an LNA unit with a modified nucleobase. Modified nucleobases or nucleosidic bases desirably base-pair with adenine, guanine, cytosine, uracil, or thymine. Exemplary oligomers contain 2 to 100, 5 to 100, 4 to 50, 5 to 50, 5 to 30, or 8 to 15 nucleic acid units. In some embodiments, one or more LNA units with natural nucleobases are incorporated into the  
10 oligonucleotide at a distance from the LNA unit having a modified base of 1 to 6 (e.g., 1 to 4) bases. In certain embodiments, at least two LNA units with natural nucleobases are flanking an LNA unit having a modified base. Desirably, at least two LNA units independently are positioned at a distance from the LNA unit having the modified base of 1 to 6 (e.g., 1 to 4 bases).

15 Desirable modified nucleobases or nucleosidic bases for use in nucleic acid compositions of the invention include optionally substituted carbon alicyclic or carbocyclic aryl groups (i.e., only carbon ring members), particularly multi-ring carbocyclic aryl groups such as groups having 2, 3, 4, 5, 6, 7, or 8 linked, particularly fused carbocyclic aryl moieties. Optionally substituted pyrene is also desirable. Such nucleobases or nucleosidic bases can  
20 provide significant performance results, as demonstrated in the examples which follow. Heteroalicyclic and heteroaromatic nucleobases or nucleosidic bases also are suitable. In some embodiments, the carbocyclic moiety is linked to the 1'-position of the LNA unit through a linker (e.g., a branched or straight alkylene or alkenylene).

Desirable LNA units have a carbon or hetero alicyclic ring with four to six ring  
25 members, e.g. a furanose ring, or other alicyclic ring structures such as a cyclopentyl, cycloheptyl, tetrahydropyranyl, oxepanyl, tetrahydrothiophenyl, pyrrolidinyl, thianyl, thiepanyl, piperidinyl, and the like. In one aspect, at least one ring atom of the carbon or hetero alicyclic group is taken to form a further cyclic linkage to thereby provide a multi-cyclic group. The cyclic linkage may include one or more, typically two atoms, of the carbon or hetero alicyclic  
30 group. The cyclic linkage also may include one or more atoms that are substituents, but not ring members, of the carbon or hetero alicyclic group. Other desirable LNA units are compounds having a substituent on the 2'-position of the central sugar moiety (e.g., ribose or xylose), or derivatives thereof, which favors the C3'-endo conformation, commonly referred to as the North (or simply N for short) conformation. These LNA units include ENA (2'-O,4'-C-

ethylene-bridged nucleic acids such as those disclosed in WO 00/47599) units as well as non-bridged riboses such as 2'-F or 2'-O-methyl.

For any of the above aspects, an exemplary control nucleic acid has  $\beta$ -D-2-deoxyribose instead of one or more bicyclic or sugar groups of a LNA unit or other modified or non-naturally-occurring units in a nucleic acid of the first population. In some embodiments, the nucleic acid or population of the invention and the control nucleic acid or population only have naturally-occurring nucleobases. If a nucleic acid in the nucleic acid or population of the invention has one or more non-naturally-occurring nucleobases, the capture efficiency of the corresponding control nucleic acid is calculated as the average capture efficiency for all of the nucleic acids that have either A, T, C, G or mC (methyl Cytosine) in each position corresponding to a non-naturally-occurring nucleobase in the nucleic acid in the first population.

#### *Complex of Target Nucleic Acids and Nucleic Acid Probes*

In one aspect, the invention features a complex of one or more target nucleic acids and nucleic acid of the invention (e.g., nucleic acid probes) in which one or more target nucleic acids are hybridized to a plurality of nucleic acids of the invention. Desirably, at least 2, 3, 4, 5, 6, 7, 10, 15, 20, 30, or 40 different target nucleic acids are hybridized. In some embodiments, the target nucleic acids are cDNA molecules reverse transcribed from a patient sample or cRNA molecules amplified from a patient sample using a T7 RNA polymerase-based linear amplification system or the like. The target nucleic acids are labeled prior to hybridization to the nucleic acids of invention.

#### *Methods for Detecting or Amplifying Target Nucleic Acids*

In one aspect, the invention features a method for detecting the presence of one or more target nucleic acids in a sample. This method involves incubating a nucleic acid sample with one or more nucleic acids of the invention under conditions that allow at least one target nucleic acid to hybridize to at least one of the nucleic acids of the invention. Desirably, hybridization is detected for at least 2, 3, 4, 5, 6, 8, 10, or 12 target nucleic acids. In some embodiments, the method further includes contacting the target nucleic acid with a second nucleic acid or a population of second nucleic acids that binds to a different region of the target molecule than the first nucleic acid. Desirably, the method further involves identifying one or more hybridized target nucleic acids and/or determining the amount of one or more hybridized target nucleic acids. In desirable embodiments, the method further includes determining the presence or absence of an mRNA splice variant of interest in the sample and/or determining the presence or absence of a mutation, deletion, and/or duplication of an exon of interest. In some

embodiments, the mutation, deletion, and/or duplication is indicative of a disease, disorder, or condition, such as cancer.

In desirable embodiments of any of the above detection methods, at least 5, 10, 15, 20, 30, 40, 50, 80, 100, 150, 200, or more target nucleic acids hybridize to the 5 nucleic acids of the invention. Desirably, the method is repeated under one or more different incubation conditions. In particular embodiments, the method is repeated at 1, 3, 5, 8, 10, 15, 20, 30, 40 or more different temperatures, cation concentrations (e.g., concentrations of monovalent cations such as  $\text{Na}^+$  and  $\text{K}^+$  or divalent cations such as  $\text{Mg}^{2+}$  and  $\text{Ca}^{2+}$ ), denaturants (e.g., hydrogen bond donors or acceptors that interfere with 10 the hydrogen bonds keeping the base-pairs together such as formamide or urea). Desirably, the method also includes identifying the target nucleic acid hybridized to the nucleic acids of the invention and/or determining the amount of the target nucleic acid hybridized to the nucleic acids of the invention. In particular embodiments, the target nucleic acids are labeled with a fluorescent group. In certain embodiments, the labeling 15 is repeated using different fluorescent groups (e.g., labelling for so-called dye-swap labeling experiments). In desirable embodiments, the determination of the amount of bound target nucleic acid involves one or more of the following: (i) adjusting for the varying intensity of the excitation light source used for detection of the hybridization, (ii) adjusting for photobleaching of the fluorescent group, and/or (iii) comparing the 20 fluorescent intensity of the target nucleic acid(s) hybridized to the nucleic acids of the invention of nucleic acids to the fluorescent intensity of a different sample of nucleic acids hybridized to the nucleic acids of the invention (e.g., a different sample hybridized to the same population of nucleic acids of the invention on the same or a different solid support such as the same chip or a different chip). Desirably, this comparison in 25 fluorescent intensity involves adjusting for a difference in the amount of the nucleic acids of the invention used for hybridization to each sample and/or adjusting for a difference in the buffer (e.g., a difference in  $\text{Mg}^{2+}$  concentration) used for hybridization to each sample or scaling for different labeling efficiencies with different fluorochromes.

30 Desirably, the target nucleic acids are cDNA molecules reverse transcribed from a patient sample or cRNA molecules amplified using a T7 RNA polymerase-based linear amplification system or the like from a patient sample. In particular embodiments, the sample has nucleic acids that are amplified using one or more primers specific for an exon of a target nucleic acid, and the method involves determining the 35 presence or absence of an mRNA splice variant with the exon in the sample. Desirably,



one or more of the primers are specific for an exon or exon-exon junction of a pathogen of interest, and the method involves determining the presence or absence of a nucleic acid with the exon in the sample.

In a desirable embodiment, the nucleic acids of the invention are covalently bonded to a solid support by reaction of a nucleoside phosphoramidite with an activated solid support, and subsequent reaction of a nucleoside phosphoramidite with an activated nucleotide or nucleic acid bound to the solid support. In some embodiments, the solid support or the growing nucleic acid bound to the solid support is activated by illumination, a photogenerated acid, or electric current.

10 In another aspect, the invention features a method for amplifying a target nucleic acid molecule. The method involves (a) incubating a first nucleic acid of the invention with a target nucleic acid under conditions that allow the first nucleic acid to bind the target nucleic acid; and (b) extending the first nucleic acid with the target nucleic acid as a template. Desirably, the method further involves contacting the target nucleic acid with a second nucleic acid (e.g., a  
15 second nucleic acid of the invention) that binds to a different region of the target nucleic acid than the first nucleic acid. In various embodiments, the sequence of the target nucleic acid is known or unknown.

In one aspect, the invention features a method of detecting a nucleic acid of a pathogen (e.g., a nucleic acid in a sample such as a blood or urine sample from a mammal). This method  
20 involves contacting a nucleic acid probe of the invention (e.g., a probe specific for an exon or a mRNA from a particular pathogen or family of pathogens) with a nucleic acid sample under conditions that allow the probe to hybridize to at least one nucleic acid in the sample. The probe is desirably at least 60, 70, 80, 90, 95, or 100% complementary to a nucleic acid of a pathogen (e.g., a bacteria, virus, or yeast such as any of the pathogens described herein).  
25 Hybridization between the probe and a nucleic acid in the sample is detected, indicating that the sample contains the corresponding nucleic acid from a pathogen. In some embodiments, the method is used to determine what strain of a pathogen has infected a mammal (e.g., a human) by determining whether a particular nucleic acid is present in the sample. In other embodiments, the probe has a universal base in a position corresponding to a nucleotide that  
30 varies among different strains of a pathogen, and thus the probe detects the presence of a nucleic acid from any of a multiple of pathogenic strains.

#### *Methods for Classifying Nucleic Acids Samples*

In one aspect, the invention features a method for classifying a test nucleic acid sample including target nucleic acids. This method involves (a) incubating a test  
35 nucleic acid sample with a one or more nucleic acids of the invention under conditions

that allow at least one of the nucleic acids in the test sample to hybridize to at least one nucleic acid of the invention, (b) detecting a hybridization pattern of the test nucleic acid sample, and (c) comparing the hybridization pattern to a hybridization pattern of a first nucleic acid standard, whereby the comparison indicates whether or not the test sample has the same classification as the first standard. Desirably, the method also includes comparing a hybridization pattern of the test nucleic acid sample to a hybridization pattern of a second standard. In various embodiments, a hybridization pattern of the test nucleic acid sample is compared to at least 3, 4, 5, 8, 10, 15, 20, 30, 40, or more standards.

10 Desirably, the method also includes identifying the hybridized target nucleic acid and/or determining the amount of hybridized target nucleic acid. In particular embodiments, the target nucleic acids are labeled with a fluorescent group. Desirably, the first nucleic acid standard is labeled with a different fluorescent group. The fluorescence of the target nucleic acids and the first nucleic acid standard can be  
15 detected simultaneously or sequentially.

In desirable embodiments, the method further includes determining the presence or absence of an mRNA splice variant of interest in the sample and/or determining the presence or absence of a mutation, deletion, and/or duplication of an exon of interest. In some embodiments, the mutation, deletion, and/or duplication is indicative of a disease,  
20 disorder, or condition, such as cancer.

In desirable embodiments, the determination of the amount of bound target nucleic acid involves one or more of the following: (i) adjusting for the varying intensity of the excitation light source used for detection of the hybridization, (ii) adjusting for photobleaching of the fluorescent group, and/or (iii) comparing the fluorescent intensity  
25 of the target nucleic acid(s) hybridized to the nucleic acids of the invention to the fluorescent intensity of a different sample of nucleic acids hybridized to the nucleic acids of the invention (e.g., a different sample hybridized to same set of nucleic acids of the invention on the same or a different solid support such as the same chip or a different chip). Desirably, this comparison in fluorescent intensity involves adjusting  
30 for a difference in the amount of the plurality used for hybridization to each sample and/or adjusting for a difference in the buffer (e.g., a difference in  $Mg^{2+}$  concentration) used for hybridization to each sample.

Desirably, the nucleic acids in the population together hybridize to at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, or 100% of the exons of a target nucleic acid. In  
35 desirable embodiments, the population includes nucleic acids that together hybridize to

at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, or 100% of the nucleic acids expressed by a particular cell or tissue. In some embodiments, the population includes nucleic acids that together hybridize to at least one exon from at least 1, 5, 10, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100% of the nucleic acid sequences expressed by a particular cell or tissue 5 at a given point in time (e.g., an expression array with sequences corresponding to the sequences of mRNA molecules expressed by a particular cell type or a cell under a particular set of conditions). Desirably, the method further includes using a nucleic acid or a region of a nucleic acid that is present in a first test sample but not present in a first standard or not present in a second test sample as a probe or primer for the detection, 10 amplification, or characterization of the nucleic acid.

In desirable embodiments of any of the above methods, at least 5, 10, 15, 20, 30, 40, 50, 80, 100, 150, 200, or more target nucleic acids hybridize to the nucleic acids of the invention. Desirably, the method is repeated under one or more different incubation or hybridization conditions. In particular embodiments, the method is repeated at 1, 3, 15 5, 8, 10, 15, 20, 30, 40 or more different temperatures, cation concentrations (e.g., concentration of monovalent cations such as  $\text{Na}^+$  and  $\text{K}^+$  or divalent cations such as  $\text{Mg}^{2+}$  and  $\text{Ca}^{2+}$ ), denaturants (e.g., hydrogen bond donors or acceptors that interfere with the hydrogen bonds keeping the base-pairs together such as formamide or urea).

In particular embodiments, the sample has nucleic acids that are amplified using 20 one or more primers specific for an exon of a target nucleic acid, and the method involves determining the presence or absence of an mRNA splice variant with the exon in the sample. Desirably, one or more of the primers are specific for an exon or exon-exon junction of a pathogen of interest, and the method involves determining the presence or absence of a nucleic acid with the exon in the sample.

25 Desirably, the comparison of the hybridization pattern of a patient nucleic acid sample to that of one or more standards is used to determine whether or not a patient has a particular disease, disorder, condition, or infection or an increased risk for a particular disease, disorder, condition, or infection. In some embodiments, the comparison is used to determine what pathogen has infected a patient and to select a therapeutic for the 30 treatment of the patient. Desirably, the comparison is used to select a therapeutic for the treatment or prevention of a disease or disorder in the patient. In yet other embodiments, the comparison is used to include or exclude the patient from a group in a clinical trial. Desirably, the comparison is used to compare the expression of nucleic acids (e.g., mRNA splice forms associated with toxicity) in the presence and absence of a 35 candidate compound (e.g., a lead compound for drug development). In other

embodiments, the comparison is used to determine differences in expression of nucleic acids (e.g., mRNA splice variants) under particular conditions (e.g., under different environmental stress conditions) or at different developmental time points. In particular embodiments, the expression of one or more members from a particular enzyme class 5 (e.g., protein kinase splice variants) is measured.

In a desirable embodiment, the nucleic acids of the invention are covalently bonded to a solid support by reaction of a nucleoside phosphoramidite with an activated solid support, and subsequent reaction of a nucleoside phosphoramidite with an activated nucleotide or nucleic acid bound to the solid support. In some embodiments, the solid support or the growing nucleic acid 10 bound to the solid support is activated by illumination, a photogenerated acid, or electric current.

The use of a variety of different monomers in the nucleic acids of the invention offers a means to "fine tune" the chemical, physical, biological, pharmacokinetic, and pharmacological properties of the nucleic acids thereby facilitating improvement in their safety and efficacy 15 profiles when used as a therapeutic drug.

#### *Applications for the Nucleic Acids of the Invention*

In another aspect, the invention features the use of one or more nucleic acids of the invention for the detection, amplification, or classification of a nucleic acid of interest or a population of nucleic acids of interest.

20 In another aspect, the invention features the use of one or more nucleic acids of the invention for alternative mRNA splice variant detection, expression profiling, comparative genomic hybridization, or real-time PCR. In exemplary real-time PCR applications, the nucleic acids are used to determine the amount of one or more target nucleic acids (e.g., mRNA splice variants) in a sample. In particular embodiments, 25 fluorescently labeled RT-PCR products from the amplification of a test nucleic acid sample are hybridized to a population of nucleic acids of the invention. Desirably, the amount of one or more RT-PCR products is measured to determine the amount of the corresponding nucleic acid in the initial sample.

In yet another aspect, the invention features the use of a nucleic acid of the invention as a 30 PCR primer or FISH probe.

#### *Methods for Selecting a Population of Nucleic Acid*

In one aspect, the invention features a method of selecting a nucleic acid for a population of nucleic acids. This method involves (a) determining the melting temperature of a nucleic acid of the invention, determining the ability of the nucleic acid 35 to self-anneal, determining the ability of the nucleic acid to hybridize to one or more

exons or introns of a target nucleic acid, and/or determining the ability of the nucleic acid to hybridize to a non-target nucleic acid, and (b) selecting the nucleic acid for inclusion or exclusion from the population based on the determination in step (a). In desirable embodiments, step (a) is performed for at least 2, 3, 4, 5, 6, 10, 20, 50, 100, 5 200, 500, 1,000, 5,000 or more nucleic acids, and a subset of the nucleic acids are selected for inclusion in the population based on the determination in step (a). Desirably, the nucleic acids with the highest melting temperatures and/or ability to hybridize to one or more exons or introns of a target nucleic acid are selected. Desirably, the nucleic acids with the lowest ability to self-anneal and/or hybridize to a 10 non-target nucleic acid are selected.

*Databases with Hybridization Patterns of Nucleic Acids Samples and/or Standards*

The invention also features a variety of databases. These databases are useful for storing the information obtained in any of the methods of the invention. These databases may also be used in the diagnosis of disease or an increased risk for a disease or in the selection of a 15 desirable therapeutic for a particular patient or class of patients.

Accordingly, in one such aspect, the invention provides an electronic database including at least 1, 10,  $10^2$ ,  $10^3$ ,  $5 \times 10^3$ ,  $10^4$ ,  $10^5$ ,  $10^6$ ,  $10^7$ ,  $10^8$ , or  $10^9$  records of a nucleic acid of interest or a population of nucleic acids of interest (e.g., one or more nucleic acids in a standard or in a test nucleic acid sample) correlated to records of its hybridization pattern to a plurality of 20 nucleic acids of the invention under one or more incubation conditions (e.g., one or more temperatures, denaturant concentrations, or salt concentrations).

In another aspect, the invention features computer including the database of the above aspect and a user interface (i) capable of displaying a hybridization pattern for a nucleic acid of interest or a population of nucleic acids of interest whose record is stored in the computer or (ii) 25 capable of displaying a nucleic acid of interest (e.g., displaying the polynucleotide sequence or another identifying characteristic of the nucleic acid of interest) or a population of nucleic acids of interest that produces a hybridization pattern whose record is stored in the computer.

*Methods for Silencing a Target Nucleic Acid in a Cell or Animal*

One method for inhibiting specific gene expression involves the use of antisense or 30 double stranded oligonucleotides which are complementary to a specific target messenger RNA (mRNA) sequence, such as a specific mRNA splice variant. Of special interest are oligonucleotides with a modified backbone (such as LNA or phosphorothioate) that are not readily degraded by endonucleases in the target cells.

In one aspect, the invention features the use of a nucleic acid of the invention for the manufacture of a pharmaceutical composition for treatment of a disease curable by an antisense or RNAi technology.

In one aspect, the invention provides a method for inhibiting the expression of a target nucleic acid in a cell. The method involves introducing into the cell a nucleic acid of the invention in an amount sufficient to specifically attenuate expression of the target nucleic acid. The introduced nucleic acid has a nucleotide sequence that is essentially complementary to a region of desirably at least 20 nucleotides of the target nucleic acid. Desirably, the cell is in a mammal.

10 In a related aspect, the invention provides a method for preventing, stabilizing, or treating a disease, disorder, or condition associated with a target nucleic acid in a mammal. This method involves introducing into the mammal a nucleic acid of the invention in an amount sufficient to specifically attenuate expression of the target nucleic acid, wherein the introduced nucleic acid has a nucleotide sequence that is essentially complementary to a region of  
15 desirably at least 20 nucleotides of the target nucleic acid.

In another aspect, the invention provides a method for preventing, stabilizing, or treating a pathogenic infection in a mammal by introducing into the mammal a nucleic acid of the invention in an amount sufficient to specifically attenuate expression of a target nucleic acid of a pathogen. The introduced nucleic acid has a nucleotide sequence that is essentially  
20 complementary to a region of desirably at least 20 nucleotides of the target nucleic acid.

In desirable embodiments of the therapeutic methods of the above aspects, the mammal is a human. In some embodiments, the introduced nucleic acid is single stranded or double stranded.

With respect to the therapeutic methods of the invention, it is not intended that the  
25 administration of nucleic acids to a mammal be limited to a particular mode of administration, dosage, or frequency of dosing; the present invention contemplates all modes of administration, including oral, intraperitoneal, intramuscular, intravenous, intraarticular, intralesional, subcutaneous, or any other route sufficient to provide a dose adequate to prevent or treat a disease (e.g., a disease associated with the expression of a target nucleic acid that is silenced  
30 with a nucleic acid of the invention). One or more nucleic acids may be administered to the mammal in a single dose or multiple doses. When multiple doses are administered, the doses may be separated from one another by, for example, one week, one month, one year, or ten years. It is to be understood that, for any particular subject, specific dosage regimes should be adjusted over time according to the individual need and the professional judgment of the person  
35 administering or supervising the administration of the compositions.

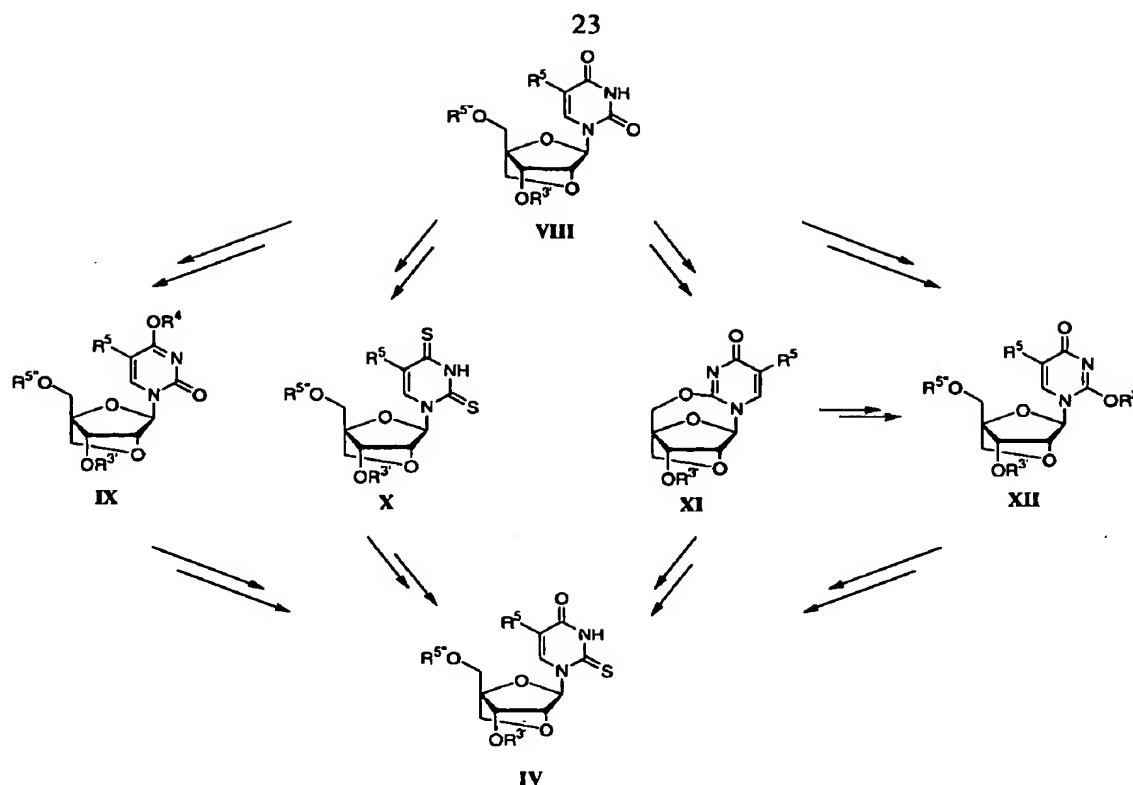
Exemplary mammals that can be treated using the methods of the invention include humans, primates such as monkeys, animals of veterinary interest (e.g., cows, sheep, goats, buffalos, and horses), and domestic pets (e.g., dogs and cats). Exemplary cells in which one or more target genes can be silenced using the methods of the invention include invertebrate, 5 plant, bacteria, yeast, and vertebrate (e.g., mammalian or human) cells.

Optimum dosages for gene silencing applications may vary depending on the relative potency of individual oligonucleotides, and can generally be estimated based on  $EC_{50}$  values found to be effective in *in vitro* and *in vivo* animal models. In general, dosage is from 0.001 ug to 100 g per kg of body weight (e.g., 0.001 ug/kg to 1 g/kg), and may be given once or more 10 daily, weekly, monthly or yearly, or even once every 2 to 20 years (U.S.P.N. 6,440,739). Persons of ordinary skill in the art can easily estimate repetition rates for dosing based on measured residence times and concentrations of the drug in bodily fluids or tissues. Following successful treatment, it may be desirable to have the patient undergo maintenance therapy to prevent the recurrence of the disease state, wherein the oligonucleotide is administered in 15 maintenance doses, ranging from 0.001 ug to 100 g per kg of body weight (e.g., 0.001 ug/kg to 1 g/kg), once or more daily, to once every 20 years. If desired, conventional treatments may be used in combination with the nucleic acids of the present invention.

Suitable carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The composition can be adapted for the mode of 20 administration and can be in the form of, for example, a pill, tablet, capsule, spray, powder, or liquid. In some embodiments, the pharmaceutical composition contains one or more pharmaceutically acceptable additives suitable for the selected route and mode of administration. These compositions may be administered by, without limitation, any parenteral route including intravenous, intra-arterial, intramuscular, subcutaneous, intradermal, 25 intraperitoneal, intrathecal, as well as topically, orally, and by mucosal routes of delivery such as intranasal, inhalation, rectal, vaginal, buccal, and sublingual. In some embodiments, the pharmaceutical compositions of the invention are prepared for administration to vertebrate (e.g., mammalian) subjects in the form of liquids, including sterile, non-pyrogenic liquids for injection, emulsions, powders, aerosols, tablets, capsules, enteric coated tablets, or 30 suppositories.

#### *Exemplary Oligomers of the Invention and Methods for Synthesizing Them*

In desirable embodiments, the invention features a method of synthesizing a nucleic acid. This method involves synthesizing a 2-thio-uridine nucleoside or nucleotide of formula IV using a compound of formula VIII, IX, X, XI, or XII as shown below. The nucleoside, 35 nucleoside phosphoramidite, or nucleotide is incorporated into a nucleic acid of the invention.



In a particular embodiment, nucleobase thiolation is performed on the O2 position of compound **XI** to form compound **IV**. In another embodiment, sulphurization on both 5 O2 and O4 in compound **VIII** generates a 2,4-dithio-uridine nucleoside or nucleotide of formula **X** which is converted into compound **IV**. In yet another embodiment, a cyclic ether of formula **XI** is transferred into compound **IV** or a 2-O-alkyl-uridine nucleoside or nucleotide of formula **XII** through reaction with the 5' position. In other embodiments, a 2-O-alkyl-uridine nucleoside or nucleotide of formula **XII** is generated 10 by direct alkylation of a uridine nucleoside or nucleotide of formula **VIII**.

In desirable embodiments  $R^4$  and  $R^2$  are each independently alkyl (e.g., methyl or ethyl), acyl (e.g., acetyl or benzoyl), or any appropriate protecting group such as silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl).  $R^{5''}$  is any appropriate protecting group such as silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, trityl(triphenylmethyl), 15 acetyl, benzoyl, or benzyl. In desirable embodiments,  $R^5$  is hydrogen, alkyl (e.g., methyl or ethyl), 1-propynyl, thiazol-2-yl, pyridine-2-yl, thien-2-yl, imidazol-2-yl, (4/5-methyl)-thiazol-2-yl, 3-(iodoacetamido)propyl, 4-[*N,N*-bis(3-aminopropyl)amino]butyl, or halo (e.g., chloro, bromo, iodo, fluoro).

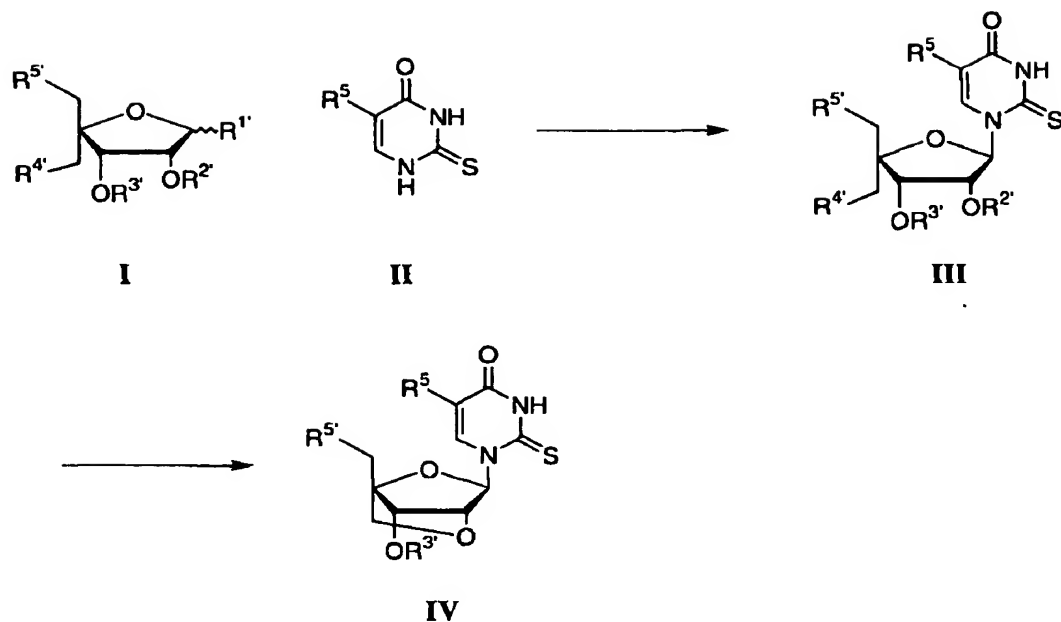
The group  $-OR^{3'}$  in the formulas **IV**, **VIII**, **IX**, **X**, **XI**, and **XII** is any of the groups listed 20 for  $R^3$  or  $R^{3'}$  in formula Ia or formula Ib or listed for  $R^3$  or  $R^{3*}$  in formula IIa, Scheme A, or



Scheme B, or the group  $-OR^3$  or  $R^3$  in the formulas IV, VIII, IX, X, XI, and XII is selected from the group consisting of H,  $-OH$ ,  $P(O(CH_2)_2CN)N(iPr)_2$ ,  $P(O(CH_2)_2CN)N(iPr)_2$ , phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (e.g., chloro, fluoro, 5 iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g., methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g., acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxy carbonyl, aralkoxycarbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, alkylthio, arylthio, 10 heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

The group  $-OR^5$  in the formulas IV, and VIII, IX, X, and XII is any of the groups listed 15 for  $R^5$  or  $R^5$  in formula Ia or formula Ib or listed for  $R^5$  or  $R^{5'}$  in formula IIa, Scheme A, or Scheme B, or the group  $-OR^5$  or  $R^5$  in the formulas IV, and VIII, IX, X, and XII is selected from the group consisting of H,  $-OH$ ,  $P(O(CH_2)_2CN)N(iPr)_2$ ,  $P(O(CH_2)_2CN)N(iPr)_2$ , phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (e.g., chloro, fluoro, 20 iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g., methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g., acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxy carbonyl, aralkoxycarbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, alkylthio, arylthio, 25 heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

In yet another aspect, the invention features a method of synthesizing a nucleic acid. 30 This method involves synthesizing a 2-thiopyrimidine nucleoside or nucleotide of formula IV using a compound of formula III or compounds of the formula I, II, and III as shown below. The nucleoside, nucleoside phosphoramidite, or nucleotide is incorporated into a nucleic acid of the invention.



In some embodiments, lewis acid-catalyzed condensation of a substituted sugar of formula **I** and a substituted 2-thio-uracil of formula **II** results in a substituted 2-thio-uridine nucleoside or nucleotide of the formula **III**. In some embodiments, a compound of formula **III** is converted 5 into a LNA 2-thiouridine nucleoside or nucleotide of formula **IV**.

In desirable embodiments  $R^4$  and  $R^5$  are, e.g., methanesulfonyloxy, p-toluenesulfonyloxy, or any appropriate protecting group such as silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, trityl(triphenylmethyl), acetyl, benzoyl, or benzyl,  $R^1$  is, e.g., acetyl, benzoyl, alkoxy (e.g., methoxy).  $R^2$  is, e.g., acetyl or benzoyl, and  $R^3$  is any appropriate 10 protecting group such as silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, trityl(triphenylmethyl), acetyl, or benzoyl. In desirable embodiments,  $R^5$  is hydrogen, alkyl (e.g. methyl or ethyl), 1-propynyl, thiazol-2-yl, pyridine-2-yl, thien-2-yl, imidazol-2-yl, (4/5-methyl)-thiazol-2-yl, 3-(iodoacetamido)propyl, 4-[*N,N*-bis(3-aminopropyl)amino]butyl, or halo (e.g., chloro, bromo, iodo, fluoro).

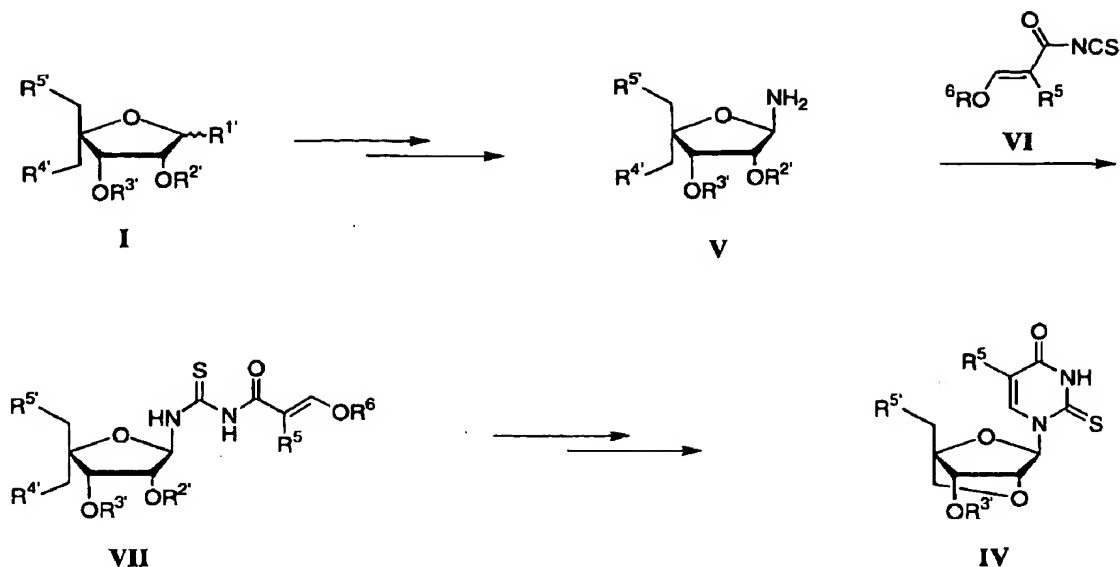
15 The group  $-OR^3$  in the formulas **I**, **III**, and **IV** is any of the groups listed for  $R^3$  or  $R^{3*}$  in formula Ia or formula Ib or listed for  $R^3$  or  $R^{3*}$  in formula IIa, Scheme A, or Scheme B, or the group  $-OR^3$  or  $R^3$  in the formulas **I**, **III**, and **IV** is selected from the group consisting of H,  $-OH$ ,  $P(O(CH_2)_2CN)N(iPr)_2$ , phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo 20 (e.g., chloro, fluoro, iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g. methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g. acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxy carbonyl, aralkoxy carbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl,

heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone 5 such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

The group  $R^5$  in the formulas I, III, and IV is any of the groups listed for  $R^5$  or  $R^{5*}$  in formula Ia or formula Ib or listed for  $R^5$  or  $R^{5*}$  in formula IIa, Scheme A, or Scheme B, or  $R^5$  in the formulas I, III, and IV is selected from the group consisting of H, -OH,

10  $P(O(CH_2)_2CN)N(iPr)_2$ , phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (e.g., chloro, fluoro, iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g., methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g. acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxy carbonyl, aralkoxy carbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, 15 heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

20 In still another aspect, the invention features a method of synthesizing a nucleic acid. This method involves synthesizing a 2-thiopyrimidine nucleoside or nucleotide of formula IV using a compound of formula VII, compounds of the formula V, VI, and VII, or compounds of the formula I, V, VI, and VII as shown below. The nucleoside, nucleoside phosphoramidite, or nucleotide is incorporated into a nucleic acid of the invention.



In some embodiments, a 2-thio-uridine nucleoside or nucleotide of the formula IV is synthesized through ring-synthesis of the nucleobase by reaction of an amino sugar of the formula V and a substituted isothiocyanate of the formula VI.

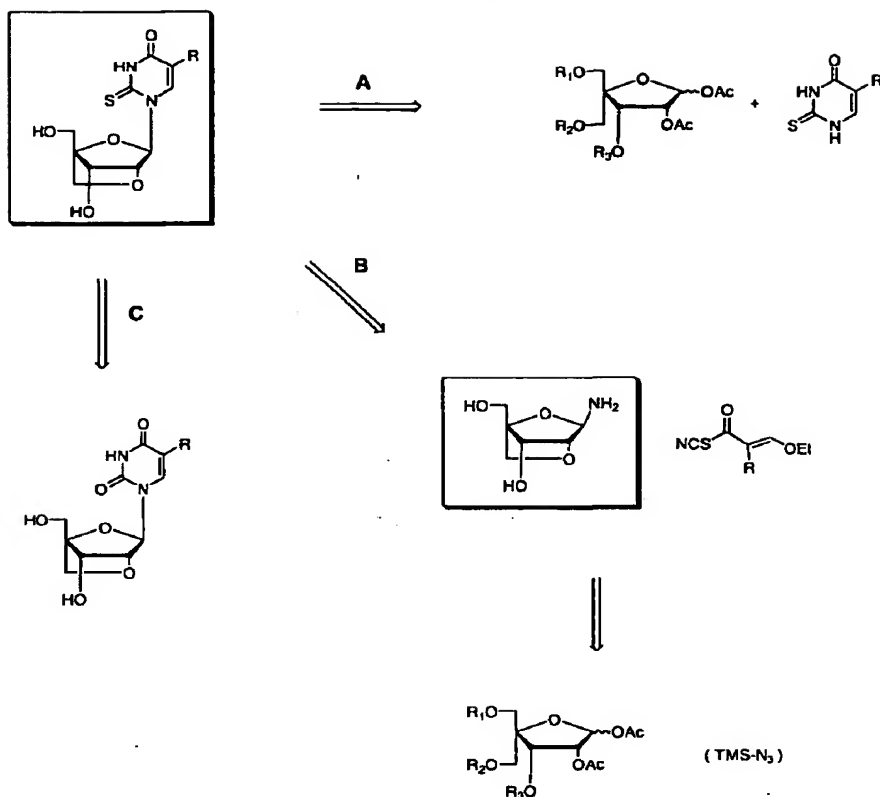
In desirable embodiments,  $R^{4'}$  and  $R^{5'}$  are each independently, e.g., methanesulfonyloxy, p-toluenesulfonyloxy, or any appropriate protecting group such as silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, trityl(triphenylmethyl), acetyl, benzoyl, or benzyl.  $R^{1'}$  is, e.g., acetyl or benzoyl or alkoxy (e.g., methoxy), and  $R^{2'}$  is, e.g., acetyl or benzoyl,  $R^{3'}$  is any appropriate protecting group such as silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, trityl(triphenylmethyl), acetyl, or benzoyl.  $R^5$  and  $R^6$  are each independently, e.g., hydrogen or alkyl (e.g. methyl or ethyl).  $R^6$  can also be, e.g., an appropriate protecting group such as silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl). In desirable embodiments,  $R^5$  is hydrogen or methyl, and  $R^6$  is methyl or ethyl.

15 The group  $-OR^{3'}$  in the formulas I, V, VII, and IV is any of the groups listed for  $R^3$  or  $R^{3'}$  in formula Ia or formula Ib or listed for  $R^3$  or  $R^{3'}$  in formula IIa, Scheme A, or Scheme B, or the group  $-OR^{3'}$  or  $R^{3'}$  in the formulas I, V, VII, and IV is selected from the group consisting of H, -OH,  $P(O(CH_2)_2CN)N(iPr)_2$ , phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (e.g., chloro, fluoro, iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g. methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g. acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxycarbonyl, aralkoxycarbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl,

alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio,amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

$R^5$  in the formulas I, V, VII, and IV is any of the groups listed for  $R^5$  or  $R^{5'}$  in formula Ia or formula Ib or listed for  $R^5$  or  $R^{5'}$  in formula IIa, Scheme A, or Scheme B, or  $R^5$  in the formulas I, V, VII, and IV is selected from the group consisting of H, -OH,  $P(O(CH_2)_2CN)N(iPr)_2$ , phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (e.g., chloro, fluoro, iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g., methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g. acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxy carbonyl, aralkoxycarbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio,amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

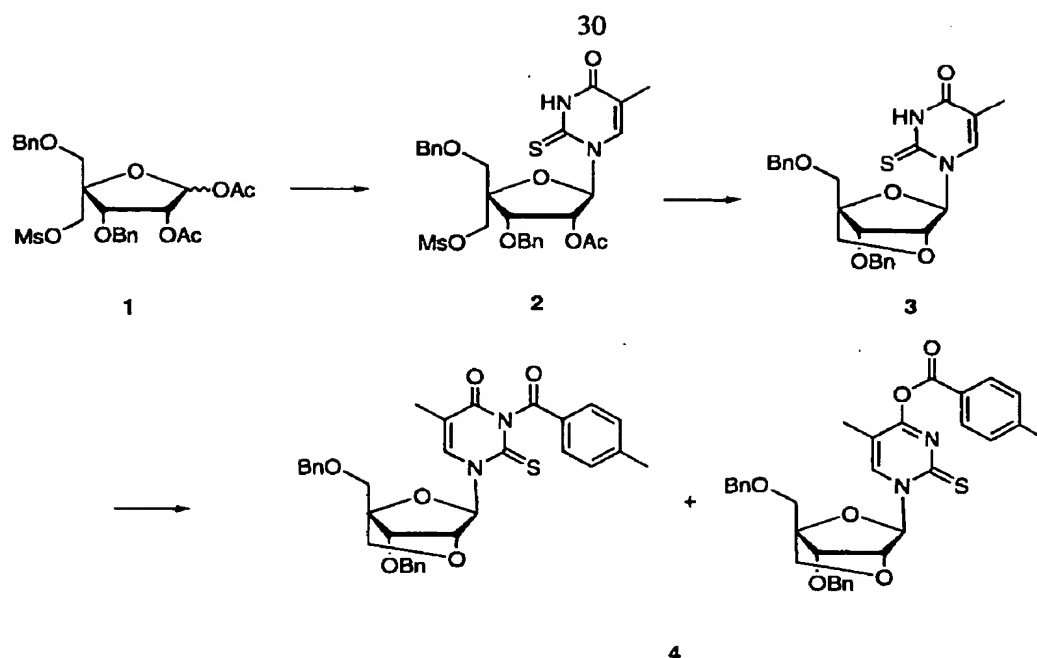
In another aspect, the invention features a method of synthesizing a nucleic acid. This method involves synthesizing a 2-thiopyrimidine nucleoside as shown below. In desirable embodiments, the method further comprises reacting one or both compounds of the formula 4 with a phosphodiamidite (e.g., 2-cyanoethyl tetraisopropylphosphorodiamidite) to produce the corresponding nucleoside phosphoramidite. The nucleoside, nucleoside phosphoramidite, or nucleotide is incorporated into a nucleic acid of the invention.



In some embodiments, a glycosyl-donor is coupled to a nucleobase as shown in pathway A. In other embodiments, ring synthesis of the nucleobase is performed as shown in pathway B. In still other embodiments, LNA-T diol is modified as shown in pathway C.

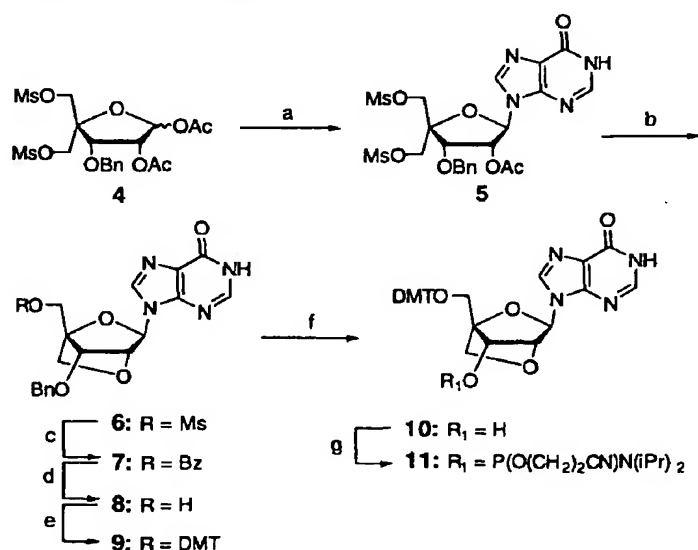
In desirable embodiments, R is hydrogen, methyl, 1-propynyl, thiazol-2-yl, pyridine-2-yl, thien-2-yl, imidazol-2-yl, (4/5-methyl)-thiazol-2-yl, 3-(iodoacetamido)propyl, 4-[*N,N*-bis(3-aminopropyl)amino]butyl, or halo (e.g., chloro, bromo, iodo, fluoro). Desirably, R<sub>1</sub>, R<sub>2</sub>, and R<sub>3</sub> are each any appropriate protecting group such as acetyl, benzyl, silyl, 4,4'-dimethoxytrityl, 10 monomethoxytrityl, or trityl(triphenylmethyl).

In another aspect, the invention features a method of synthesizing a nucleic acid. This method involves synthesizing a 2-thiopyrimidine nucleoside or nucleotide of formula 4 using a compound of formula 3, compounds of the formula 2 and 3, or compounds of the formula 1, 2, 3, and 4 as shown below. The nucleoside, nucleoside phosphoramidite, or nucleotide is 15 incorporated into a nucleic acid of the invention. This method can also be performed using any other appropriate protecting groups instead of Bn (benzyl), Ac (acetyl), or Ms (methanesulfonyl).



In desirable embodiments, the method further comprises reacting one or both compounds of the formula 4 with a phosphodiamidite (e.g., 2-cyanoethyl tetraisopropylphosphorodiamidite) to produce the corresponding nucleoside phosphoramidite.

5 In another aspect, the invention features a method of synthesizing a nucleic acid. This method involves synthesizing a nucleoside or nucleotide of formula 10 or 11 using a compound of any one of the formula 6-9, compounds of the formula 5 and any one of the formulas 6-9, or compounds of the formula 4, 5, and any one of the formulas 6-9 as shown below. The nucleoside, nucleoside phosphoramidite, or nucleotide is incorporated into a nucleic acid of the invention. This method can also be performed using any other appropriate protecting groups instead of DMT, Bn, Ac, or Ms.



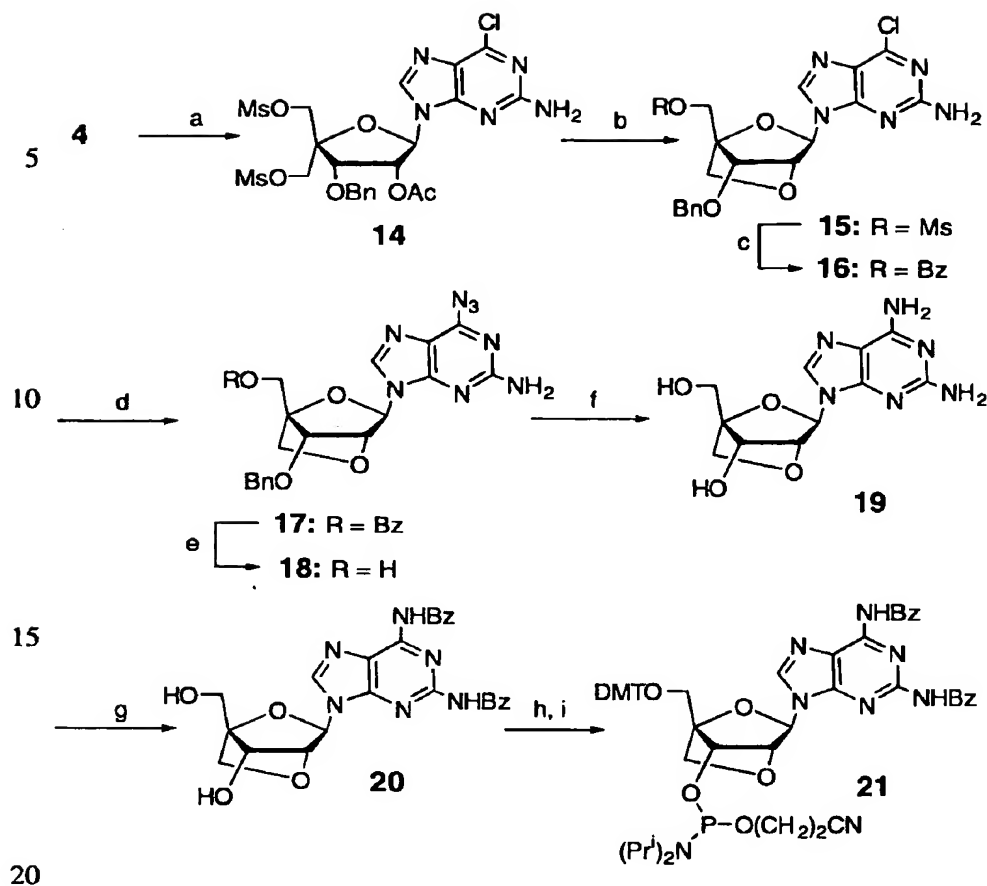
In some embodiments, a compound of formula 4 is used as a glycosyl donor in a coupling reaction with silylated hypoxanthine to form a compound of the formula 5. In certain embodiments, a compound of the formula 5 is used in a ring closing reaction to form a compound of the formula 6. Desirably, deprotection of the 5'-hydroxy group of compound 6 is performed by displacing the 5'-O-mesyl group with sodium benzoate to produce a compound of the formula 7 that is converted into a compound of the formula 8 after saponification of the 5'-benzoate. In some embodiments, compound 8 is converted to a DMT-protected compound 9 prior to debenzoylation of the 3'-O-hydroxy group. In desirable embodiments, a phosphoramidite of the formula 11 is generated by phosphorylation of a nucleoside of the formula 10.

In desirable embodiments, the  $R_1$  is H or  $P(O(CH_2)_2CN)N(iPr)_2$ . In other embodiments, the group  $R_1$  or  $-OR_1$  is any of the groups listed for  $R^3$  or  $R^{3*}$  in formula Ia or formula Ib or listed for  $R^3$  or  $R^{3*}$  in formula IIa, Scheme A, or Scheme B, or the group

$-OR_1$  or  $R_1$  is selected from the group consisting of -OH,  $P(O(CH_2)_2CN)N(iPr)_2$ , phosphate, 15 phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (e.g., chloro, fluoro, iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g., methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g., acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, 20 aryloxy carbonyl, aralkoxy carbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroaryl sulfonyl, alkylsulfinyl, arylsulfinyl, heteroaryl sulfinyl, alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone 25 such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

In another aspect, the invention features a method of synthesizing a nucleic acid. This method involves synthesizing a nucleoside or nucleotide of formula 20 or 21 as shown below, in which compound 4 is the same sugar shown in the above aspect. The nucleoside, nucleoside phosphoramidite, or nucleotide is incorporated into a nucleic acid of the invention. This 30 method can also be performed using any other appropriate protecting groups instead of DMT, Bn, Bz (benzoyl), Ac, or Ms. Additionally, the method can be performed with any other halogen (e.g., fluoro or bromo) instead of chloro.





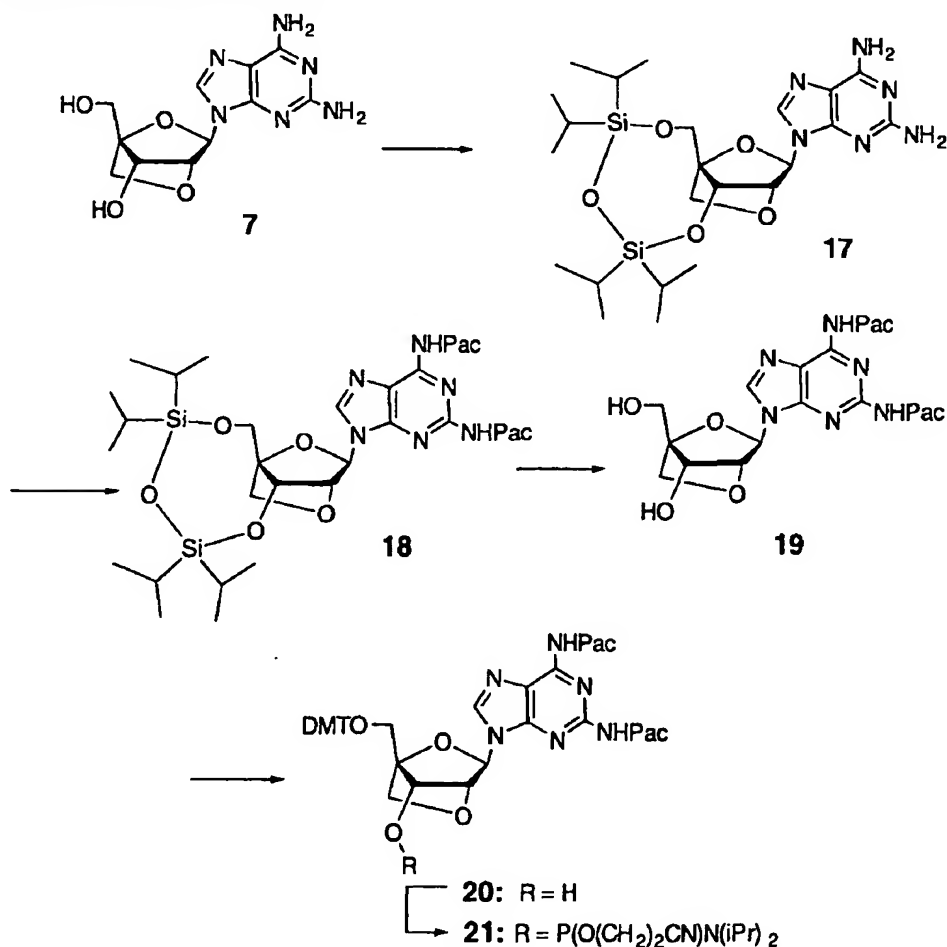
In desirable embodiments to promote the ring closing reaction, a solution of compound **14** in aqueous 1,4-dioxane is treated with sodium hydroxide to give a bicyclic compound **15**. In some embodiments, sodium benzoate is used for displacement of 5'-mesylate of compound **15** to give compound **16**. In some embodiments, compound **17** is formed by reaction of compound **16** with sodium azide. In some embodiments, compound **18** is produced by saponification of the 5'-benzoate of compound **17**. In certain embodiments, hydrogenation of compound **18** produces compound **19**. In certain embodiments, the peracelation method is used to benzoylate the 2- and 6-amino groups of compound **19**, yielding **20**, which is desirably converted into the phosphoramidite compound **21**.

In a related aspect, the invention features a derivative of a compound of the formula **20** or **21** as described in the above aspect in which 3' -OH or -OP(O(CH<sub>2</sub>)<sub>2</sub>CN)N(iPr)<sub>2</sub> group is replaced by any other group is selected from the group consisting of phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (*e.g.*, chloro, fluoro, iodo, or bromo), optionally substituted aryl, (*e.g.*,

phenyl or benzyl), alkyl (e.g. methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g. acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxycarbonyl, aralkoxycarbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, 5 alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

- 10 In yet another aspect, the invention features a method of synthesizing a nucleic acid. This method involves synthesizing a nucleoside or nucleotide of formula 20 or 21 as shown below. The nucleoside, nucleoside phosphoramidite, or nucleotide is incorporated into a nucleic acid of the invention. This method can also be performed using any other appropriate protecting groups instead of DMT.

15

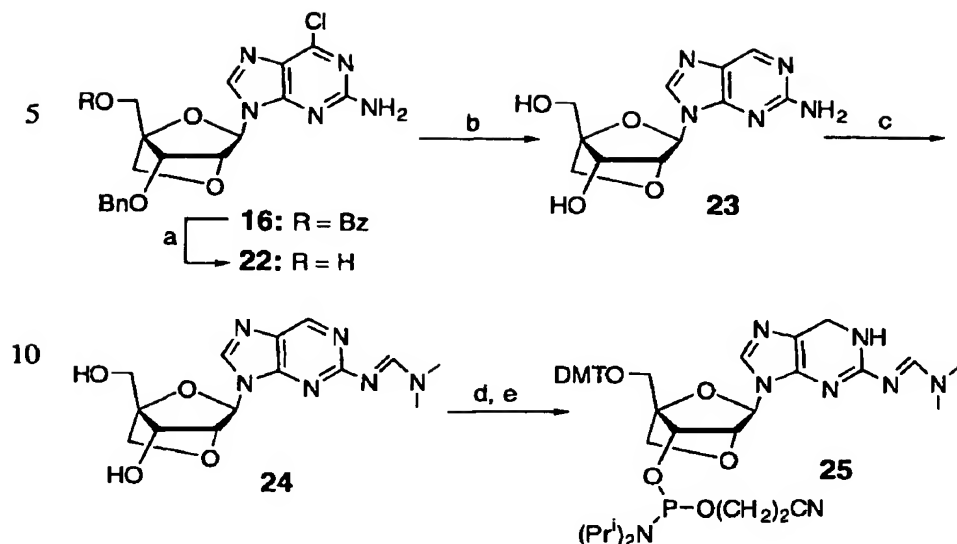


In some embodiments, compound 17 is formed by reaction of compound 7 with 1,3-dichloro-1,1,3,3-tetraisopropylidisiloxane. Desirably, compound 18 is formed by reaction of compound 17 with phenoxyacetic anhydride. In some embodiments, compound 19 is generated by reaction of compound 18 with acid. Desirably, compound 20 is produced by reacting compound 19 with DMT-Cl. In desirably embodiments, compound 20 is reacted with 2-cyanoethyl tetraisopropylphosphorodiamidite to give the phosphoramidite 21.

In desirable embodiments, the R is H or  $\text{P}(\text{O}(\text{CH}_2)_2\text{CN})\text{N}(\text{iPr})_2$ . In other embodiments, the R or -OR is any of the groups listed for  $\text{R}^3$  or  $\text{R}^{3'}$  in formula Ia or formula Ib or listed for  $\text{R}^3$  or  $\text{R}^{3'}$  in formula IIa, Scheme A, or Scheme B, or the group

10 -OR or R is selected from the group consisting of -OH,  $\text{P}(\text{O}(\text{CH}_2)_2\text{CN})\text{N}(\text{iPr})_2$ , phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (e.g., chloro, fluoro, iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g., methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g., acetyl or benzoyl), aroyl, aralkyl, hydroxy, 15 hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxy carbonyl, aralkoxycarbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or 20 trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

In yet another aspect, the invention features a method of synthesizing a nucleic acid. This method involves synthesizing a nucleoside or nucleotide of formula 24 or 25 as shown below. The nucleoside, nucleoside phosphoramidite, or nucleotide is incorporated into a 25 nucleic acid of the invention. This method can also be performed using any other appropriate protecting groups instead of Bz, Bn, and DMT. Additionally, the method can be performed with any other halogen (e.g., fluoro or bromo) instead of chloro.



15

In some embodiments, the compound **16** is formed from compounds **4**, **14**, and **15** as illustrated in an aspect above. Desirably, the 5'-O-benzoyl group of compound **16** is hydrolyzed by aqueous sodium hydroxide to give compound **22**. Compound **23** is desirably produced by incubation of compound **22** in the presence of palladium hydroxide and ammonium formate. Desirably, the 2-amine of compound **23** is selectively protected with an amidine group after treatment with *N,N*-dimethylformamide dimethyl acetal to yield compound **24**. In some embodiments, the diol **24** is 5'-O-DMT protected and 3'-O-phosphitylated produce the phosphoramidite LNA-2AP compound **25**.

In some embodiments, compound **25** has one of the following groups instead of the  $\text{P}(\text{O}(\text{CH}_2)_2\text{CN})\text{N}(\text{iPr})_2$  group: any of the groups listed for  $\text{R}^3$  or  $\text{R}^{3'}$  in formula Ia or formula Ib or listed for  $\text{R}^3$  or  $\text{R}^{3'}$  in formula IIa, Scheme A, or Scheme B, or a group selected from the group consisting of -OH, phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, methyl phosphonate, halo (e.g., chloro, fluoro, iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g., methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g. acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxy carbonyl, aralkoxycarbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl,

monomethoxytrityl), or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

In another aspect, the invention features a nucleic acid of the invention that includes a compound of the formula **6pC** or the product of a compound of the formula **6pC** treated with ammonia as described herein. In a related aspect, the invention features a method of synthesizing a nucleic acid that involves performing one or more of the steps described herein for the synthesis of a compound of the formula **6pC** or the product of a compound of the formula **6pC** treated with ammonia.

10 In yet another aspect, the invention features a method of synthesizing a nucleic acid. This method involves one or more of any of the nucleosides or nucleotides of the invention with (i) any other nucleoside or nucleotide of the invention, (ii) any other nucleoside or nucleotide of formula Ia, formula Ib, formula IIa, Scheme A, or Scheme B, and/or (iii) any naturally-occurring nucleoside or nucleotide. Desirably, the method involves reacting one or more  
15 nucleoside phosphoramidites of any of the above aspects with a nucleotide or nucleic acid.

*Methods for Synthesis of Nucleic Acids on a Solid Support*

In another aspect, the invention provides a method for the synthesis of a population of nucleic acids (e.g., a population of nucleic acids of the invention) on a solid support. This method involves the reaction of a plurality of nucleoside  
20 phosphoramidites with an activated solid support (e.g., a solid support with an activated linker) and the subsequent reaction of a plurality of nucleoside phosphoramidites with activated nucleotides or nucleic acids bound to the solid support.

In some embodiments of any of the above aspects, the solid support or the growing nucleic acid bound to the solid support is activated by illumination, a  
25 photogenerated acid, or electric current. In desirable embodiments, one or more spots or regions (e.g., a region with an area of less than 1 cm<sup>2</sup>, 0.1 cm<sup>2</sup>, 0.01 cm<sup>2</sup>, 1 mm<sup>2</sup>, or 0.1 mm<sup>2</sup> that desirably contains one particular nucleic acid monomer or oligomer) on the solid support are irradiated to produce a photogenerated acid that removes the 5'-OH protecting group of one or more nucleic acid monomers or oligomers to which a  
30 nucleotide is subsequently added. In other embodiments, an electric current is applied to one or more spots or regions (e.g., a region with an area of less than 1 cm<sup>2</sup>, 0.1 cm<sup>2</sup>, 0.01 cm<sup>2</sup>, 1 mm<sup>2</sup>, or 0.1 mm<sup>2</sup> that desirably contains one particular nucleic acid monomer or oligomer) on the solid support to remove an electrochemically sensitive protecting group of one or more nucleic acid monomers or oligomers to which a  
35 nucleotide is subsequently added. In still other embodiments, one or more spots or

regions (e.g., a region with an area of less than 1 cm<sup>2</sup>, 0.1 cm<sup>2</sup>, 0.01 cm<sup>2</sup>, 1 mm<sup>2</sup>, or 0.1 mm<sup>2</sup> that desirably contains one particular nucleic acid monomer or oligomer) on the solid support are irradiated to remove a photosensitive protecting group of one or more nucleic acid monomers or oligomers to which a nucleotide is subsequently added. In various embodiments, the solid support (e.g., chip, coverslip, microscope glass slide, quartz, or silicon) is less than 1, 0.5, 0.1, or 0.05 mm thick.

#### *Methods for the Synthesis of Nucleic Acids*

In another aspect, the invention features a method of reacting a population of nucleic acids of the invention with one or more nucleic acids. This method involves incubating an immobilized population of nucleic acids of the invention with a solution that includes one or more probes (e.g., at least 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 80, 100, or 150 different nucleic acids) and one or more target nucleic acids (e.g., at least 2, 3, 4, 5, 10, 15, 20, 30, 40, 50, 60, 80, 100, or 150 different target nucleic acids). The incubation is performed in the presence of a ligase under conditions that allow the ligase to covalently react one or more immobilized nucleic acids with one or more nucleic acid probes in solution that hybridize to the same target nucleic acid. Desirably, at least 2, 5, 10, 15, 20, 30, 40, 50, 80, or 100 pairs of immobilized nucleic acids and nucleic acid probes are ligated. In various embodiments, the incubation occurs between 15 and 45°C, such as between 20 and 40°C or between 25 and 35°C.

#### *Desirable Embodiments of Any of the Aspects of the Invention*

In other embodiments of any of various aspects of the invention, a nucleic acid probe or primer specifically hybridizes to a target nucleic acid but does not substantially hybridize to non-target molecules, which include other nucleic acids in a cell or biological sample having a sequence that is less than 99, 95, 90, 80, or 70% identical or complementary to that of the target nucleic acid. Desirably, the amount of the these non-target molecules hybridized to, or associated with, the nucleic acid probe or primer, as measured using standard assays, is 2-fold, desirably 5-fold, more desirably 10-fold, and most desirably 50-fold lower than the amount of the target nucleic acid hybridized to, or associated with, the nucleic acid probe or primer. In other embodiments, the amount of a target nucleic acid hybridized to, or associated with, the nucleic acid probe or primer, as measured using standard assays, is 2-fold, desirably 5-fold, more desirably 10-fold, and most desirably 50-fold greater than the amount of a control nucleic acid hybridized to, or associated with, the nucleic acid probe or primer. In certain embodiments, the nucleic acid probe or primer RNA is substantially complementary (e.g., at least 80, 90, 95, 98, or 100% complementary) to a target nucleic acid or a group of target nucleic acids from a cell. In other embodiments, the probe or primer is homologous to multiple

RNA or DNA molecules, such as RNA or DNA molecules from the same gene family. In other embodiments, the probe or primer is homologous to a large number of RNA or DNA molecules. In desirable embodiments, the probe or primer binds to nucleic acids which have polynucleotide sequences that differ in sequence at a position that corresponds to the position of a universal base in the probe or primer. Examples of control nucleic acids include nucleic acids with a random sequence or nucleic acids known to have little, if any, affinity for the nucleic acid probe or primer. In some embodiments, the target nucleic acid is an RNA, DNA, or cDNA molecule.

Desirably, the association constant ( $K_a$ ) of the nucleic acid toward a complementary target molecule is higher than the association constant of the complementary strands of the double stranded target molecule. In some desirable embodiments, the melting temperature of a duplex between the nucleic acid and a complementary target molecule is higher than the melting temperature of the complementary strands of the double stranded target molecule.

In some embodiments, the LNA-pyrene is in a position corresponding to the position of a non-base (e.g., a unit without a base) in another nucleic acid, such as a target nucleic acid. Incorporation of pyrene in a DNA strand that is hybridized against the four natural bases decreases the  $T_m$  by  $-4.5^\circ\text{C}$  to  $-6.8^\circ\text{C}$ ; however, incorporation of pyrene in a DNA strand in a position opposite a non-base only decreases the  $T_m$  by  $-2.3^\circ\text{C}$  to  $-4.6^\circ\text{C}$ , most likely due to the better accommodation of the pyrene in the B-type duplex (Matray and Kool, J. Am. Chem. Soc. 120, 6191, 1998). Thus, incorporation of LNA-pyrene into a nucleic acid in a position opposite a non-base (e.g., a unit without a base or a unit with a small group such as a noncyclic group instead of a base) in a target nucleic acid may also minimize any potential decrease in  $T_m$  due to the pyrene substitution.

In various embodiments, the number of molecules in the plurality of nucleic acids of the invention is at least 2, 4, 5, 6, 7, 8, or 10-fold greater than the number of molecules in the test nucleic acid sample. In some embodiments, a LNA is a triplex-forming oligonucleotide.

In desirable embodiments of any of the aspects of the invention, the target nucleic acids (e.g., cDNA molecules reverse transcribed from a patient sample) are fragmented using an enzyme such as a uracil-DNA glycosylase (e.g., *E. coli* uracil-DNA glycosylase) or using chemical hydrolysis such as alkaline hydrolysis. In various embodiments, the average size of the fragmented nucleic acids is between 300 and 50 nucleic acids, such as approximately 300, 200, 100, or 50 nucleotides.

#### *Advantages*

The present invention has a variety of advantages related to nucleic acid analysis methods. The ability to equalize melting temperatures of a series of nucleic acids is

generally applicable and desirable in all situations where more than one sequence is used simultaneously (e.g. DNA arrays with more than one capture probe, PCR and especially multiplex PCR, homogeneous assays such as Taqman and Molecular beacon). Sample preparation of specific sequences (e.g., DNA or RNA extraction using capture probes on filters or magnetic beads) is another area where melting temperature equalization of specific probe sequences is useful.

For example, the invention provides high affinity nucleotides (e.g., LNA and other high affinity nucleotides with a modified base and/or backbone) that can be used, e.g., arrays of the invention. In particular, the nucleic acids of the invention containing LNA units exhibited a surprising ability to discriminate between different mRNA splice variants compared to naturally-occurring nucleic acids. If desired, universal bases can be added as part of flanking regions in capture probes (e.g., probes of an array) to stabilize hybridization with high affinity nucleotides in the capture probes. Replacement of one or more DNA-t nucleotides with LNA-T and/or replacement of one or more DNA-a nucleotides with LNA-A reduces the variability of melting temperatures for capture probes of similar length but different GC and AT content by desirably at least 10, 20, 30, 40 or 50%. Additionally, replacement of one or more DNA-t nucleotides with LNA-T and/or replacement of one or more DNA-c with LNA-C increases the stability of a large number of capture probes, while desirably avoiding self-complementary sequences with LNA:LNA base-pairs within a capture probe that would otherwise reduce or eliminate the binding of target molecules to the probe. Although a general T and C substitution may not reduce the variability of melting temperatures of the probes, this substitution increases the melting temperature and binding efficiency of many capture probes that contain these two nucleotides.

The invention also provides a general substitution algorithm for enhancement of the hybridization signal of a test nucleic acid sample by inclusion of high affinity monomers (e.g., LNA and other high affinity nucleotides with a modified base and/or backbone) in the array. This method increases the stability and binding affinity of capture probes while avoiding substitutions in positions that may form self-complementary base-pairs which may otherwise inhibit binding to a target molecule. The substitution algorithm is broadly useful for specialized arrays, as well as for PCR primers and FISH probes.

Other features and advantages of the invention will be apparent from the following detailed description.

#### *Definitions*

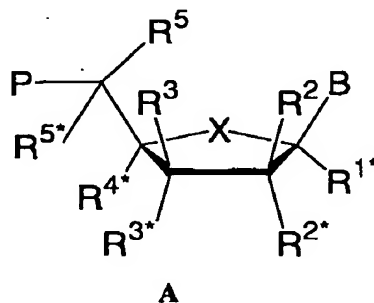
When used herein, the term "LNA" (Locked Nucleoside Analogues) refers to nucleoside analogues (e.g., bicyclic nucleoside analogues, e.g., as disclosed in WO 9914226) either



incorporated in an oligonucleotide or as a discrete chemical species (e.g., LNA nucleoside and LNA nucleotide). The term "monomeric LNA" may, e.g., refer to the monomers LNA A, LNA T, LNA C, or any other LNA monomers.

By "LNA unit" is meant an individual LNA monomer (e.g., an LNA nucleoside or LNA nucleotide) or an oligomer (e.g., an oligonucleotide or nucleic acid) that includes at least one LNA monomer. LNA units as disclosed in WO 99/14226 are in general particularly desirable modified nucleic acids for incorporation into an oligonucleotide of the invention. Additionally, the nucleic acids may be modified at either the 3' and/or 5' end by any type of modification known in the art. For example, either or both ends may be capped with a protecting group, 10 attached to a flexible linking group, attached to a reactive group to aid in attachment to the substrate surface, etc. Desirable LNA units and their method of synthesis also are disclosed in WO 0056746, WO 0056748, WO 0066604, Morita *et al.*, Bioorg. Med. Chem. Lett. 12(1):73-76, 2002; Hakansson *et al.*, Bioorg. Med. Chem. Lett. 11(7):935-938, 2001; Koshkin *et al.*, J. Org. Chem. 66(25):8504-8512, 2001; Kvaerno *et al.*, J. Org. Chem. 66(16):5498-5503, 2001; 15 Hakansson *et al.*, J. Org. Chem. 65(17):5161-5166, 2000; Kvaerno *et al.*, J. Org. Chem. 65(17):5167-5176, 2000; Pfundheller *et al.*, Nucleosides Nucleotides 18(9):2017-2030, 1999; and Kumar *et al.*, Bioorg. Med. Chem. Lett. 8(16):2219-2222, 1998.

By "LNA modified oligonucleotide" is meant a oligonucleotide comprising at least one LNA monomeric unit of the general scheme A, described *infra*, having the below described 20 illustrative examples of modifications:



wherein X is selected from -O-, -S-, -N(R<sup>N</sup>)-, -C(R<sup>6</sup>R<sup>6\*</sup>)-, -O-C(R<sup>7</sup>R<sup>7\*</sup>)-, -C(R<sup>6</sup>R<sup>6\*</sup>)-O-, -S- 25 C(R<sup>7</sup>R<sup>7\*</sup>)-, -C(R<sup>6</sup>R<sup>6\*</sup>)-S-, -N(R<sup>N\*</sup>)-C(R<sup>7</sup>R<sup>7\*</sup>)-, -C(R<sup>6</sup>R<sup>6\*</sup>)-N(R<sup>N</sup>)-, and -C(R<sup>6</sup>R<sup>6\*</sup>)-C(R<sup>7</sup>R<sup>7\*</sup>).

B is selected from a modified base as discussed above e.g. an optionally substituted carbocyclic aryl such as optionally substituted pyrene or optionally substituted pyrenylmethylglycerol, or an optionally substituted heteroalicyclic or optionally substituted heteroaromatic such as optionally substituted pyridyloxazole, optionally substituted pyrrole, 30 optionally substituted diazole or optionally substituted triazole moieties; hydrogen, hydroxy,

optionally substituted C<sub>1-4</sub>-alkoxy, optionally substituted C<sub>1-4</sub>-alkyl, optionally substituted C<sub>1-4</sub>-acyloxy, nucleobases, DNA intercalators, photochemically active groups, thermochemically active groups, chelating groups, reporter groups, and ligands.

P designates the radical position for an internucleoside linkage to a succeeding 5 monomer, or a 5'-terminal group, such internucleoside linkage or 5'-terminal group optionally including the substituent R<sup>5</sup>. One of the substituents R<sup>2</sup>, R<sup>2\*</sup>, R<sup>3</sup>, and R<sup>3\*</sup> is a group P\* which designates an internucleoside linkage to a preceding monomer, or a 2'/3'-terminal group. The substituents of R<sup>1\*</sup>, R<sup>4\*</sup>, R<sup>5</sup>, R<sup>5\*</sup>, R<sup>6</sup>, R<sup>6\*</sup>, R<sup>7</sup>, R<sup>7\*</sup>, R<sup>N</sup>, and the ones of R<sup>2</sup>, R<sup>2\*</sup>, R<sup>3</sup>, and R<sup>3\*</sup> not designating P\* each designates a biradical comprising about 1-8 groups/atoms selected from -10 C(R<sup>a</sup>R<sup>b</sup>)-, -C(R<sup>a</sup>)=C(R<sup>a</sup>)-, -C(R<sup>a</sup>)=N-, -C(R<sup>a</sup>)-O-, -O-, -Si(R<sup>a</sup>)<sub>2</sub>-, -C(R<sup>a</sup>)-S-, -S-, -SO<sub>2</sub>-, -C(R<sup>a</sup>)-N(R<sup>b</sup>)-, -N(R<sup>a</sup>)-, and >C=Q, wherein Q is selected from -O-, -S-, and -N(R<sup>a</sup>)-, and R<sup>a</sup> and R<sup>b</sup> each is independently selected from hydrogen, optionally substituted C<sub>1-12</sub>-alkyl, optionally substituted C<sub>2-12</sub>-alkenyl, optionally substituted C<sub>2-12</sub>-alkynyl, hydroxy, C<sub>1-12</sub>-alkoxy, C<sub>2-12</sub>-alkenyloxy, carboxy, C<sub>1-12</sub>-alkoxycarbonyl, C<sub>1-12</sub>-alkylcarbonyl, formyl, aryl, aryloxy-carbonyl, 15 aryloxy, arylcarbonyl, heteroaryl, hetero-aryloxy-carbonyl, heteroaryloxy, heteroarylcarbonyl, amino, mono- and di(C<sub>1-6</sub>-alkyl)amino, carbamoyl, mono- and di(C<sub>1-6</sub>-alkyl)-amino-carbonyl, amino-C<sub>1-6</sub>-alkyl-aminocarbonyl, mono- and di(C<sub>1-6</sub>-alkyl)amino-C<sub>1-6</sub>-alkyl-aminocarbonyl, C<sub>1-6</sub>-alkyl-carbonylamino, carbamido, C<sub>1-6</sub>-alkanoyloxy, sulphono, C<sub>1-6</sub>-alkylsulphonyloxy, nitro, azido, sulphanyl, C<sub>1-6</sub>-alkylthio, halogen, DNA intercalators, photochemically active 20 groups, thermochemically active groups, chelating groups, reporter groups, and ligands, where aryl and heteroaryl may be optionally substituted, and where two geminal substituents R<sup>a</sup> and R<sup>b</sup> together may designate optionally substituted methylene (=CH<sub>2</sub>), and wherein two non-geminal or geminal substituents selected from R<sup>a</sup>, R<sup>b</sup>, and any of the substituents R<sup>1\*</sup>, R<sup>2</sup>, R<sup>2\*</sup>, R<sup>3</sup>, R<sup>3\*</sup>, R<sup>4\*</sup>, R<sup>5</sup>, R<sup>5\*</sup>, R<sup>6</sup> and R<sup>6\*</sup>, R<sup>7</sup>, and R<sup>7\*</sup> which are present and not involved in P, P\* or the 25 biradical(s) together may form an associated biradical selected from biradicals of the same kind as defined before; the pair(s) of non-geminal substituents thereby forming a mono- or bicyclic entity together with (i) the atoms to which said non-geminal substituents are bound and (ii) any intervening atoms.

Each of the substituents R<sup>1\*</sup>, R<sup>2</sup>, R<sup>2\*</sup>, R<sup>3</sup>, R<sup>4\*</sup>, R<sup>5</sup>, R<sup>5\*</sup>, R<sup>6</sup> and R<sup>6\*</sup>, R<sup>7</sup>, and R<sup>7\*</sup> which are 30 present and not involved in P, P\* or the biradical(s), is independently selected from hydrogen, optionally substituted C<sub>1-12</sub>-alkyl, optionally substituted C<sub>2-12</sub>-alkenyl, optionally substituted C<sub>2-12</sub>-alkynyl, hydroxy, C<sub>1-12</sub>-alkoxy, C<sub>2-12</sub>-alkenyloxy, carboxy, C<sub>1-12</sub>-alkoxycarbonyl, C<sub>1-12</sub>-alkylcarbonyl, formyl, aryl, aryloxy-carbonyl, aryloxy, arylcarbonyl, heteroaryl, heteroaryloxy-carbonyl, heteroaryloxy, heteroarylcarbonyl, amino, mono- and di(C<sub>1-6</sub>-alkyl)amino, carbamoyl, 35 mono- and di(C<sub>1-6</sub>-alkyl)-amino-carbonyl, amino-C<sub>1-6</sub>-alkyl-aminocarbonyl, mono- and di(C<sub>1-6</sub>-

alkyl)amino-C<sub>1-6</sub>-alkyl-aminocarbonyl, C<sub>1-6</sub>-alkyl-carbonylamino, carbamido, C<sub>1-6</sub>-alkanoyloxy, sulphono, C<sub>1-6</sub>-alkylsulphonyloxy, nitro, azido, sulphanyl, C<sub>1-6</sub>-alkylthio, halogen, DNA intercalators, photochemically active groups, thermochemically active groups, chelating groups, reporter groups, and ligands, where aryl and heteroaryl may be optionally substituted, and where two geminal substituents together may designate oxo, thioxo, imino, or optionally substituted methylene, or together may form a spiro biradical consisting of a 1-5 carbon atom(s) alkylene chain which is optionally interrupted and/or terminated by one or more heteroatoms/groups selected from -O-, -S-, and -(NR<sup>N</sup>)- where R<sup>N</sup> is selected from hydrogen and C<sub>1-4</sub>-alkyl, and where two adjacent (non-geminal) substituents may designate an additional bond resulting in a double bond; and R<sup>N\*</sup>, when present and not involved in a biradical, is selected from hydrogen and C<sub>1-4</sub>-alkyl; and basic salts and acid addition salts thereof.

Exemplary 5', 3', and/or 2' terminal groups include -H, -OH, halo (e.g., chloro, fluoro, iodo, or bromo), optionally substituted aryl, (e.g., phenyl or benzyl), alkyl (e.g., methyl or ethyl), alkoxy (e.g., methoxy), acyl (e.g., acetyl or benzoyl), aroyl, aralkyl, hydroxy, hydroxyalkyl, alkoxy, aryloxy, aralkoxy, nitro, cyano, carboxy, alkoxycarbonyl, aryloxy carbonyl, aralkoxy carbonyl, acylamino, aroylamine, alkylsulfonyl, arylsulfonyl, heteroarylsulfonyl, alkylsulfinyl, arylsulfinyl, heteroarylsulfinyl, alkylthio, arylthio, heteroarylthio, aralkylthio, heteroaralkylthio, amidino, amino, carbamoyl, sulfamoyl, alkene, alkyne, protecting groups (e.g., silyl, 4,4'-dimethoxytrityl, monomethoxytrityl, or trityl(triphenylmethyl)), linkers (e.g., a linker containing an amine, ethylene glycol, quinone such as anthraquinone), detectable labels (e.g., radiolabels or fluorescent labels), and biotin.

It is understood that references herein to a nucleic acid unit, nucleic acid residue, LNA unit, or similar term are inclusive of both individual nucleoside units and nucleotide units and nucleoside units and nucleotide units within an oligonucleotide.

A "modified base" or other similar term refers to a composition (e.g., a non-naturally occurring nucleobase or nucleosidic base) which can pair with a natural base (e.g., adenine, guanine, cytosine, uracil, and/or thymine) and/or can pair with a non-naturally occurring nucleobase or nucleosidic base. Desirably, the modified base provides a T<sub>m</sub> differential of 15, 12, 10, 8, 6, 4, or 2°C or less as described herein. Exemplary modified bases are described in EP 1 072 679 and WO 97/12896.

By "nucleobase" is meant the naturally occurring nucleobases adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U) as well as non-naturally occurring nucleobases such as xanthine, diaminopurine, 8-oxo-N<sup>6</sup>-methyladenine, 7-deazaxanthine, 7-deazaguanine, N<sup>4</sup>,N<sup>4</sup>-ethanocytosine, N<sup>6</sup>,N<sup>6</sup>-ethano-2,6-diaminopurine, 5-methylcytosine (mC), 5-(C<sup>3</sup>-C<sup>6</sup>)-alkynyl-cytosine, 5-fluorouracil, 5-bromouracil, pseudoisocytosine, 2-hydroxy-5-methyl-4-

triazolopyridin, isocytosine, isoguanine, inosine and the "non-naturally occurring" nucleobases described in Benner et al., U.S. Pat. No. 5,432,272 and Susan M. Freier and Karl-Heinz Altmann, *Nucleic Acids Research*, 1997, vol. 25, pp 4429-4443. The term "nucleobase" thus includes not only the known purine and pyrimidine heterocycles, but also heterocyclic 5 analogues and tautomers thereof. Further naturally and non-naturally occurring nucleobases include those disclosed in U.S. Pat. No. 3,687,808 (Merigan, et al.), in Chapter 15 by Sanghvi, in *Antisense Research and Application*, Ed. S. T. Crooke and B. Lebleu, CRC Press, 1993, in Englisch et al., *Angewandte Chemie, International Edition*, 1991, 30, 613-722 (see especially pages 622 and 623, and in the *Concise Encyclopedia of Polymer Science and Engineering*, J. I. Kroschwitz Ed., John Wiley & Sons, 1990, pages 858-859, Cook, *Anti-Cancer Drug Design* 1991, 6, 585-607, each of which are hereby incorporated by reference in their entirety). The term "nucleosidic base" or "base unit" is further intended to include compounds such as heterocyclic compounds that can serve like nucleobases including certain "universal bases" that are not nucleosidic bases in the most classical sense but serve as nucleosidic bases. Especially 15 mentioned as universal bases are 3-nitropyrrole, optionally substituted indoles (e.g., 5-nitroindole), and optionally substituted hypoxanthine. Other desirable universal bases include, pyrrole, diazole or triazole derivatives, including those universal bases known in the art.

As described herein, various groups of an LNA unit may be optionally substituted. A "substituted" group such as a nucleobase or nucleosidic base and the like may be substituted by 20 other than hydrogen at one or more available positions, typically 1 to 3 or 4 positions, by one or more suitable groups such as those disclosed herein. Suitable groups that may be present on a "substituted" group include e.g. halogen such as fluoro, chloro, bromo and iodo; cyano; hydroxyl; nitro; azido; alkanoyl such as a C<sub>1-6</sub> alkanoyl group such as acyl and the like; carboxamido; alkyl groups including those groups having 1 to about 12 carbon atoms, or 1, 2, 3, 25 4, 5, or 6 carbon atoms; alkenyl and alkynyl groups including groups having one or more unsaturated linkages and from 2 to 12 carbon, or 2, 3, 4, 5 or 6 carbon atoms; alkoxy groups including those having one or more oxygen linkages and from 1 to about 12 carbon atoms, or 1, 2, 3, 4, 5 or 6 carbon atoms; aryloxy such as phenoxy; alkylthio groups including those moieties having one or more thioether linkages and from 1 to about 12 carbon atoms, or 1, 2, 3, 4, 5 or 6 30 carbon atoms; alkylsulfinyl groups including those moieties having one or more sulfinyl linkages and from 1 to about 12 carbon atoms, or 1, 2, 3, 4, 5, or 6 carbon atoms; alkylsulfonyl groups including those moieties having one or more sulfonyl linkages and from 1 to about 12 carbon atoms, or 1, 2, 3, 4, 5, or 6 carbon atoms; aminoalkyl groups such as groups having one or more N atoms and from 1 to about 12 carbon atoms, or 1, 2, 3, 4, 5 or 6 carbon atoms; 35 carbocyclic aryl having 6 or more carbons; aralkyl having 1 to 3 separate or fused rings and

from 6 to about 18 carbon ring atoms, with benzyl being a desirable group; aralkoxy having 1 to 3 separate or fused rings and from 6 to about 18 carbon ring atoms, with O-benzyl being a desirable group; or a heteroaromatic or heteroalicyclic group having 1 to 3 separate or fused rings with 3 to about 8 members per ring and one or more N, O or S atoms, e.g. coumarinyl, quinolinyl, pyridyl, pyrazinyl, pyrimidyl, furyl, pyrrolyl, thienyl, thiazolyl, oxazolyl, imidazolyl, indolyl, benzofuranyl, benzothiazolyl, tetrahydrofuranyl, tetrahydropyranyl, piperidinyl, morpholino and pyrrolidinyl.

By "oxy-LNA monomer or unit" is meant any nucleoside or nucleotide which contains an oxygen atom in a 2'-4' linkage.

10 A "non-oxy-LNA" monomer or unit is broadly defined as any nucleoside or nucleotide which does not contain an oxygen atom in a 2'-4' linkage. Examples of non-oxy-LNA monomers include 2'-deoxynucleotides (DNA) or nucleotides (RNA) or any analogues of these monomers which are not oxy-LNA, such as for example the thio-LNA and amino-LNA described herein with respect to formula 1a and in Singh *et al.* J. Org. Chem. 1998, 6, 6078-9, 15 and the derivatives described in Susan M. Freier and Karl-Heinz Altmann, Nucleic Acids Research, 1997, vol 25, pp 4429-4443.

By "universal base" is meant a naturally-occurring or desirably a non-naturally occurring compound or moiety that can pair with a natural base (e.g., adenine, guanine, cytosine, uracil, and/or thymine), and that has a  $T_m$  differential of 15, 12, 10, 8, 6, 4, or 2°C or 20 less as described herein.

By "oligonucleotide," "oligomer," or "oligo" is meant a successive chain of monomers (e.g., glycosides of heterocyclic bases) connected via internucleoside linkages. The linkage between two successive monomers in the oligo consist of 2 to 4, desirably 3, groups/atoms selected from  $-\text{CH}_2-$ ,  $-\text{O}-$ ,  $-\text{S}-$ ,  $-\text{NR}^{\text{H}}-$ ,  $>\text{C}=\text{O}$ ,  $>\text{C}=\text{NR}^{\text{H}}$ ,  $>\text{C}=\text{S}$ ,  $-\text{Si}(\text{R}'')_2-$ ,  $-\text{SO}-$ ,  $-\text{S}(\text{O})_2-$ ,  $-\text{P}(\text{O})_2-$ , 25  $-\text{PO}(\text{BH}_3)-$ ,  $-\text{P}(\text{O},\text{S})-$ ,  $-\text{P}(\text{S})_2-$ ,  $-\text{PO}(\text{R}')-$ ,  $-\text{PO}(\text{OCH}_3)-$ , and  $-\text{PO}(\text{NHR}^{\text{H}})-$ , where  $\text{R}^{\text{H}}$  is selected from hydrogen and  $\text{C}_{1-4}$ -alkyl, and  $\text{R}''$  is selected from  $\text{C}_{1-6}$ -alkyl and phenyl. Illustrative examples of such linkages are  $-\text{CH}_2-\text{CH}_2-\text{CH}_2-$ ,  $-\text{CH}_2-\text{CO}-\text{CH}_2-$ ,  $-\text{CH}_2-\text{CHOH}-\text{CH}_2-$ ,  $-\text{O}-\text{CH}_2-\text{O}-$ ,  $-\text{O}-\text{CH}_2-\text{CH}_2-$ ,  $-\text{O}-\text{CH}_2-\text{CH}=\text{}$  (including  $\text{R}^{\text{S}}$  when used as a linkage to a succeeding monomer),  $-\text{CH}_2-\text{CH}_2-\text{O}-$ ,  $-\text{NR}^{\text{H}}-\text{CH}_2-\text{CH}_2-$ ,  $-\text{CH}_2-\text{CH}_2-\text{NR}^{\text{H}}-$ ,  $-\text{CH}_2-\text{NR}^{\text{H}}-\text{CH}_2-$ ,  $-\text{O}-\text{CH}_2-\text{CH}_2-\text{NR}^{\text{H}}-$ , 30  $-\text{NR}^{\text{H}}-\text{CO}-\text{O}-$ ,  $-\text{NR}^{\text{H}}-\text{CO}-\text{NR}^{\text{H}}-$ ,  $-\text{NR}^{\text{H}}-\text{CS}-\text{NR}^{\text{H}}-$ ,  $-\text{NR}^{\text{H}}-\text{C}(=\text{NR}^{\text{H}})-\text{NR}^{\text{H}}-$ ,  $-\text{NR}^{\text{H}}-\text{CO}-\text{CH}_2-\text{NR}^{\text{H}}-$ ,  $-\text{O}-\text{CO}-\text{O}-$ ,  $-\text{O}-\text{CO}-\text{CH}_2-\text{O}-$ ,  $-\text{O}-\text{CH}_2-\text{CO}-\text{O}-$ ,  $-\text{CH}_2-\text{CO}-\text{NR}^{\text{H}}-$ ,  $-\text{O}-\text{CO}-\text{NR}^{\text{H}}-$ ,  $-\text{NR}^{\text{H}}-\text{CO}-\text{CH}_2-$ ,  $-\text{O}-\text{CH}_2-\text{CO}-\text{NR}^{\text{H}}-$ ,  $-\text{O}-\text{CH}_2-\text{CH}_2-\text{NR}^{\text{H}}-$ ,  $-\text{CH}=\text{N}-\text{O}-$ ,  $-\text{CH}_2-\text{NR}^{\text{H}}-\text{O}-$ ,  $-\text{CH}_2-\text{O}-\text{N}=\text{}$  (including  $\text{R}^{\text{S}}$  when used as a linkage to a succeeding monomer),  $-\text{CH}_2-\text{O}-\text{NR}^{\text{H}}-$ ,  $-\text{CO}-\text{NR}^{\text{H}}-\text{CH}_2-$ ,  $-\text{CH}_2-\text{NR}^{\text{H}}-\text{O}-$ ,  $-\text{CH}_2-\text{NR}^{\text{H}}-\text{CO}-$ ,  $-\text{O}-\text{NR}^{\text{H}}-\text{CH}_2-$ ,  $-\text{O}-\text{NR}^{\text{H}}-$ ,  $-\text{O}-\text{CH}_2-\text{S}-$ ,  $-\text{S}-\text{CH}_2-\text{O}-$ ,  $-\text{CH}_2-\text{CH}_2-\text{S}-$ ,  $-\text{O}-\text{CH}_2-$  35  $\text{CH}_2-\text{S}-$ ,  $-\text{S}-\text{CH}_2-\text{CH}=\text{}$  (including  $\text{R}^{\text{S}}$  when used as a linkage to a succeeding monomer),  $-\text{S}-\text{CH}_2-$

CH<sub>2</sub>-, -S-CH<sub>2</sub>-CH<sub>2</sub>-O-, -S-CH<sub>2</sub>-CH<sub>2</sub>-S-, -CH<sub>2</sub>-S-CH<sub>2</sub>-, -CH<sub>2</sub>-SO-CH<sub>2</sub>-, -CH<sub>2</sub>-SO<sub>2</sub>-CH<sub>2</sub>-, -O-SO-O-, -O-S(O)<sub>2</sub>-O-, -O-S(O)<sub>2</sub>-CH<sub>2</sub>-, -O-S(O)<sub>2</sub>-NR<sup>H</sup>-, -NR<sup>H</sup>-S(O)<sub>2</sub>-CH<sub>2</sub>-, -O-S(O)<sub>2</sub>-CH<sub>2</sub>-, -O-P(O)<sub>2</sub>-O-, -O-P(O,S)-O-, -O-P(S)<sub>2</sub>-O-, -S-P(O)<sub>2</sub>-O-, -S-P(O,S)-O-, -S-P(S)<sub>2</sub>-O-, -O-P(O)<sub>2</sub>-S-, -O-P(O,S)-S-, -O-P(S)<sub>2</sub>-S-, -S-P(O)<sub>2</sub>-S-, -S-P(O,S)-S-, -S-P(S)<sub>2</sub>-S-, -O-PO(R'')-O-, -O-PO(OCH<sub>3</sub>)-O-, -O-PO(OCH<sub>2</sub>CH<sub>3</sub>)-O-, -O-PO(OCH<sub>2</sub>CH<sub>2</sub>S-R)-O-, -O-PO(BH<sub>3</sub>)-O-, -O-PO(NHR<sup>N</sup>)-O-, -O-P(O)<sub>2</sub>-NR<sup>H</sup>-, -NR<sup>H</sup>-P(O)<sub>2</sub>-O-, -O-P(O,NR<sup>H</sup>)-O-, -CH<sub>2</sub>-P(O)<sub>2</sub>-O-, -O-P(O)<sub>2</sub>-CH<sub>2</sub>-, and -O-Si(R'')<sub>2</sub>-O-; among which -CH<sub>2</sub>-CO-NR<sup>H</sup>-, -CH<sub>2</sub>-NR<sup>H</sup>-O-, -S-CH<sub>2</sub>-O-, -O-P(O)<sub>2</sub>-O-, -O-P(O,S)-O-, -O-P(S)<sub>2</sub>-O-, -NR<sup>H</sup>-P(O)<sub>2</sub>-O-, -O-P(O,NR<sup>H</sup>)-O-, -O-PO(R'')-O-, -O-PO(CH<sub>3</sub>)-O-, and -O-PO(NHR<sup>N</sup>)-O-, where R<sup>H</sup> is selected from hydrogen and C<sub>1-4</sub>-alkyl, 10 and R'' is selected from C<sub>1-6</sub>-alkyl and phenyl, are especially desirable. Further illustrative examples are given in Mesmaeker *et. al.*, Current Opinion in Structural Biology 1995, 5, 343-355 and Susan M. Freier and Karl-Heinz Altmann, Nucleic Acids Research, 1997, vol 25, pp 4429-4443. The left-hand side of the internucleoside linkage is bound to the 5-membered ring as substituent P\* at the 3'-position, whereas the right-hand side is bound to the 5'-position of a 15 preceding monomer.

By "succeeding monomer" is meant the neighboring monomer in the 5'-terminal direction, and by "preceding monomer" is meant the neighboring monomer in the 3'-terminal direction.

By "LNA spiked oligo" is meant an oligonucleotide, such as a DNA oligonucleotide, 20 wherein at least one unit (and preferably not all units) has been substituted by the corresponding LNA nucleoside monomer.

The term "T<sub>m</sub>" is used in reference to the "melting temperature." The melting temperature is the temperature at which 50% of a population of double-stranded nucleic acid molecules becomes dissociated into single strands. The equation for calculating the T<sub>m</sub> of 25 nucleic acids is well-known in the art. The T<sub>m</sub> of a hybrid nucleic acid is often estimated using a formula adopted from hybridization assays in 1 M salt, and commonly used for calculating T<sub>m</sub> for PCR primers: T<sub>m</sub> = [(number of A+T) x 2°C + (number of G+C) x 4°C]. C. R. Newton *et al.* PCR, 2nd Ed., Springer-Verlag (New York: 1997), p. 24. This formula was found to be inaccurate for primers longer than 20 nucleotides. Id. Other more sophisticated computations 30 exist in the art which take structural as well as sequence characteristics into account for the calculation of T<sub>m</sub>. A calculated T<sub>m</sub> is merely an estimate; the optimum temperature is commonly determined empirically.

A nucleic acid compound that has a T<sub>m</sub> differential of a specified amount (e.g., less than 15, 12, 10, 8, 6, 4, 2, or 1°C) means the nucleic acid exhibits that specified T<sub>m</sub> differential when

incorporated into a specified 9-mer oligonucleotide with respect to the four complementary variants, as defined immediately below.

Unless otherwise indicated, a  $T_m$  differential provided by a particular modified base is calculated by the following protocol (steps a) through d)):

5 a) incorporating the modified base of interest into the following oligonucleotide 5'-d(GTGAMATGC), wherein M is the modified base;

b) mixing  $1.5 \times 10^{-6}M$  of the oligonucleotide having incorporated therein the modified base with each of  $1.5 \times 10^{-6}M$  of the four oligonucleotides having the sequence 3'-d(CACTYTACG), wherein Y is A, C, G, T, respectively, in a buffer of 10mM sodium phosphate, 100 mM sodium chloride, 0.1 mM EDTA, pH 7.0;

c) allowing the oligonucleotides to hybridize; and

d) detecting the  $T_m$  for each of the four hybridized nucleotides by heating the hybridized nucleotides and observing the temperature at which the maximum of the first derivative of the melting curve recorded at a wavelength of 260 nm is obtained.

15 Unless otherwise indicated, a  $T_m$  differential for a particular modified base is determined by subtracting the highest  $T_m$  value determined in steps a) through d) immediately above from the lowest  $T_m$  value determined by steps a) through d) immediately above.

By "variance in  $T_m$ " is meant the variance in the values of the melting temperatures for a population of nucleic acids. The  $T_m$  for each nucleic acid is determined by experimentally  
20 measuring or computationally predicting the temperature at which 50% of a population double-stranded molecules with the sequence of the nucleic acid becomes dissociated into single strands. For a nucleic acid with only A, T, C, G, and/or U bases, the  $T_m$  is the temperature at which 50% of a population of 100% complementary double-stranded molecules with the sequence of the nucleic acid becomes dissociated into single strands. For determining the  $T_m$   
25 variance when a nucleic acid has one or more nucleobases other than A, T, C, G, or U, the  $T_m$  of this "modified" nucleic acid is approximated by determining the  $T_m$  for each possible double stranded molecule in which one strand is the modified nucleic acid and the other strand has either A, T, C, or G in each position corresponding to a nucleobase other than A, T, C, G, or U in the modified nucleic acid. For example, if the modified nucleic acid has the sequence XMX  
30 in which X is 0, 1, or more A, T, C, G, or U bases and M is any other nucleobase or nucleosidic base, the  $T_m$  is calculated for each possible double stranded molecule in which one strand is XMX and the other strand is X'YX' in which X' is the base complementary to the corresponding X base and Y is either A, T, C, or G. The average is then calculated for the  $T_m$  values for each possible double stranded molecule (i.e., four possible duplexes per modified nucleobase or

nucleoside base in the modified nucleic acid) and used as the approximate  $T_m$  value for the modified nucleic acid.

By "capture efficiency" is meant the amount of target nucleic acid(s) bound to a particular nucleic acid or a population of nucleic acids. Standard methods can be used to calculate the capture efficiency by measuring the amount of bound target nucleic acid(s) and/or measuring the amount of unbound target nucleic acid(s). The capture efficiency of a nucleic acid or nucleic acid population of the invention is typically compared to the capture efficiency of a control nucleic acid or nucleic acid population under the same incubation conditions (e.g., using same buffer and temperature).

10 For example, a control nucleic acid may have  $\beta$ -D-2-deoxyribose instead of one or more bicyclic or sugar groups of a LNA unit or other modified or non-naturally-occurring units in a nucleic acid of the invention. In some embodiments, the nucleic acid of the invention and the control nucleic acid only have naturally-occurring nucleobases. If a nucleic acid of the invention has one or more non-naturally-occurring nucleobases, the capture efficiency of the  
15 corresponding control nucleic acid is calculated as the average capture efficiency for all of the nucleic acids that have either A, T, C, or G in each position corresponding to a non-naturally-occurring nucleobase in the nucleic acid of the invention.

Monomers are referred to as being "complementary" if they contain nucleobases that can form hydrogen bonds according to Watson-Crick base-pairing rules (e.g., G with C, A with  
20 T, or A with U) or other hydrogen bonding motifs such as for example diaminopurine with T, inosine with C, and pseudocytosine with G.

By "substantially complementarity" is meant having a sequence that is at least 60, 70, 80, 90, 95, or 100% complementary to that of another sequence. Sequence complementarity is typically measured using sequence analysis software with the default parameters specified  
25 therein (e.g., Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). This software program matches similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications.

The term "homology" refers to a degree of complementarity. There can be partial  
30 homology or complete homology (i.e., identity). A partially complementary sequence that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous."

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to a probe that can hybridize to a  
35 strand of the double-stranded nucleic acid sequence under conditions of low stringency, e.g.



using a hybridization buffer comprising 20% formamide in 0.8M saline/0.08M sodium citrate (SSC) buffer at a temperature of 37°C and remaining bound when subject to washing once with that SSC buffer at 37°C.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to a probe that can hybridize to (i.e., is the complement of) the single-stranded nucleic acid template sequence under conditions of low stringency, e.g. using a hybridization buffer comprising 20% formamide in 0.8M saline/0.08M sodium citrate (SSC) buffer at a temperature of 37°C and remaining bound when subject to washing once with that SSC buffer at 37°C.

10 By "internal probe" is meant a nucleic acid (e.g., a probe or primer) that hybridizes to either only one exon or only one intron of a nucleic acid (e.g., mRNA). The internal probe may hybridize to the 5' end of the exon or intron, the 3' end of the exon or intron, or between the 5' end and the 3' end of the exon or intron. Desirably, the internal probe is at least 90, 95, 96, 97, 98, 99, or 100% identical to the corresponding region of a target nucleic acid.

15 By "merged probe" is meant a nucleic acid (e.g., a probe or primer) that hybridizes to more than one exon and/or intron of a nucleic acid (e.g., mRNA). Desirably, the merged probe hybridizes to two consecutive exons (e.g., exons in a mature mRNA transcript that may or may not be consecutive in the corresponding DNA molecule). In another desirable embodiment, the merged probe hybridizes to an exon and the consecutive intron. In desirable embodiments, the  
20 merged probe hybridizes to the same number of nucleotides in each exon or to the same number of nucleotides in the exon and intron. In various embodiments, the length of the region of the merged probe that hybridizes to one exon differs by less than 60, 40, 20, 10, or 5% from the length of the region of the merged probe that hybridizes to the other exon or to the intron. Desirably, the merged probe is at least 90, 95, 96, 97, 98, 99, or 100% identical to the  
25 corresponding region of a target nucleic acid.

By "poly-T<sub>20</sub> tail" is meant an LNA polymer consisting of 20 LNA-T units often added as a tail to a nucleic acid sequence.

By "mixmer" or "mixmer probe" is meant a nucleic acid (e.g., a probe or primer) that contains at least one LNA unit and at least one RNA or DNA unit (e.g., a naturally-occurring  
30 RNA or DNA unit).

By "corresponding unmodified reference nucleobase" is meant a nucleobase that is not part of an LNA unit and is in the same orientation as the nucleobase in an LNA unit.

By "mutation" is meant an alteration in a naturally-occurring or reference nucleic acid sequence, such as an insertion, deletion, frameshift mutation, silent mutation, nonsense

mutation, or missense mutation. Desirably, the amino acid sequence encoded by the nucleic acid sequence has at least one amino acid alteration from a naturally-occurring sequence.

By "selecting" is meant substantially partitioning a molecule from other molecules in a population. Desirably, the partitioning provides at least a 2-fold, desirably, a 30-fold, more desirably, a 100-fold, and most desirably, a 1,000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. The selection step may be repeated a number of times, and different types of selection steps may be combined in a given approach. The population desirably contains at least  $10^9$  molecules, more desirably at least  $10^{11}$ ,  $10^{13}$ , or  $10^{14}$  molecules and, most desirably, at least  $10^{15}$  molecules.

10 By a "population" is meant more than one nucleic acid. A "population" according to the invention desirably means more than  $10^1$ ,  $10^2$ ,  $10^3$ , or  $10^4$  different molecules.

By "photochemically active groups" is meant compounds which are able to undergo chemical reactions upon irradiation with light. Illustrative examples of functional groups are quinones, especially 6-methyl-1,4-naphthoquinone, anthraquinone, naphthoquinone, and 1,4-15 dimethyl-anthraquinone, diazirines, aromatic azides, benzophenones, psoralens, diazo compounds, and diazirino compounds.

By "thermochemically reactive group" is meant a functional group which is able to undergo thermochemically-induced covalent bond formation with other groups. Illustrative examples of functional parts of thermochemically reactive groups are carboxylic acids, 20 carboxylic acid esters such as activated esters, carboxylic acid halides such as acid fluorides, acid chlorides, acid bromide, acid iodides, carboxylic acid azides, carboxylic acid hydrazides, sulfonic acids, sulfonic acid esters, sulfonic acid halides, semicarbazides, thiosemicarbazides, aldehydes, ketones, primary alcohols, secondary alcohols, tertiary alcohols, phenols, alkyl halides, thiols, disulphides, primary amines, secondary amines, tertiary amines, hydrazines, 25 epoxides, maleimides, and boronic acid derivatives.

By "chelating group" is meant a molecule that contains more than one binding site and frequently binds to another molecule, atom, or ion through more than one binding site at the same time. Examples of functional parts of chelating groups are iminodiacetic acid, nitrilotriacetic acid, ethylenediamine tetraacetic acid (EDTA), and aminophosphonic acid.

30 By "reporter group" is meant a group which is detectable either by itself or as a part of an detection series. Examples of functional parts of reporter groups are biotin, digoxigenin, fluorescent groups (e.g., groups which are able to absorb electromagnetic radiation, e.g. light or X-rays, of a certain wavelength, and which subsequently reemit the energy absorbed as radiation of longer wavelength; such as dansyl (5-dimethylamino)-1-naphthalenesulfonyl), 35 DOXYL (N-oxy-4,4-dimethyloxazolidine), PROXYL (N-oxy-2,2,5,5-tetramethylpyrrolidine),

TEMPO (N-oxy-2,2,6,6-tetramethylpiperidine), dinitrophenyl, acridines, coumarins, Cy3 and Cy5 (trademarks for Biological Detection Systems, Inc.), erythrosine, coumaric acid, umbelliferone, Texas red, rhodamine, tetramethyl rhodamine, Rox, 7-nitrobenzo-2-oxa-1-diazole (NBD), pyrene, fluorescein, Europium, Ruthenium, Samarium, and other rare earth metals), radioisotopic labels, chemiluminescence labels (i.e., labels that are detectable via the emission of light during a chemical reaction), spin labels (a free radical e.g., substituted organic nitroxides) or other paramagnetic probes (e.g.,  $\text{Cu}^{2+}$  or  $\text{Mg}^{2+}$ ) bound to a biological molecule being detectable by the use of electron spin resonance spectroscopy), enzymes (such as peroxidases, alkaline phosphatases,  $\beta$ -galactosidases, and glucose oxidases), antigens, antibodies, haptens (e.g., groups which are able to combine with an antibody, but which cannot initiate an immune response by itself, such as peptides and steroid hormones), carrier systems for cell membrane penetration, fatty acid units, steroid moieties (cholesteryl), vitamin A, vitamin D, vitamin E, folic acid peptides for specific receptors, groups for mediating endocytose, epidermal growth factor (EGF), bradykinin, and platelet derived growth factor (PDGF). Especially desirable groups are biotin, fluorescein, Texas Red, rhodamine, dinitrophenyl, digoxigenin, Ruthenium, Europium, Cy5, and Cy3.

By "ligand" is meant a compound which binds. Ligands can comprise functional groups such as aromatic groups (such as benzene, pyridine, naphthalene, anthracene, and phenanthrene), heteroaromatic groups (such as thiophene, furan, tetrahydrofuran, pyridine, dioxane, and pyrimidine), carboxylic acids, carboxylic acid esters, carboxylic acid halides, carboxylic acid azides, carboxylic acid hydrazides, sulfonic acids, sulfonic acid esters, sulfonic acid halides, semicarbazides, thiosemicarbazides, aldehydes, ketones, primary alcohols, secondary alcohols, tertiary alcohols, phenols, alkyl halides, thiols, disulphides, primary amines, secondary amines, tertiary amines, hydrazines, epoxides, maleimides,  $\text{C}_1$ - $\text{C}_{20}$  alkyl groups optionally interrupted or terminated with one or more heteroatoms such as oxygen atoms, nitrogen atoms, and/or sulphur atoms, optionally containing aromatic or mono/polyunsaturated hydrocarbons, polyoxyethylene such as polyethylene glycol, oligo/polyamides such as poly- $\alpha$ -alanine, polyglycine, polylysine, peptides, oligo/polysaccharides, oligo/polyphosphates, toxins, antibiotics, cell poisons, and steroids. "Affinity ligands" include functional groups or biomolecules that have a specific affinity for sites on particular proteins, antibodies, poly- and oligosaccharides, and other biomolecules.

It should be understood that the above-mentioned specific examples under DNA intercalators, photochemically active groups, thermochemically active groups, chelating groups, reporter groups, and ligands correspond to the "active/functional" part of the groups in question. For the person skilled in the art it is furthermore clear that DNA intercalators, photochemically

active groups, thermochemically active groups, chelating groups, reporter groups, and ligands are typically represented in the form M-K- where M is the "active/functional" part of the group in question and where K is a spacer through which the "active/functional" part is attached to the 5- or 6-membered ring. Thus, it should be understood that the group B, in the case where B is selected from DNA intercalators, photochemically active groups, thermochemically active groups, chelating groups, reporter groups, and ligands, has the form M-K-, where M is the "active/functional" part of the DNA intercalator, photochemically active group, thermochemically active group, chelating group, reporter group, and ligand, respectively, and where K is an optional spacer comprising 1-50 atoms, desirably 1-30 atoms, in particular 1-15 10 atoms, between the 5- or 6-membered ring and the "active/functional" part.

By "spacer" is meant a thermochemically and photochemically non-active distance-making group and is used to join two or more different moieties of the types defined above. Spacers are selected on the basis of a variety of characteristics including their hydrophobicity, hydrophilicity, molecular flexibility and length (e.g., Hermanson *et. al.*, "Immobilized Affinity 15 Ligand Techniques," Academic Press, San Diego, California (1992). Generally, the length of the spacers is less than or about 400 Å, in some applications desirably less than 100 Å. The spacer, thus, comprises a chain of carbon atoms optionally interrupted or terminated with one or more heteroatoms, such as oxygen atoms, nitrogen atoms, and/or sulphur atoms. Thus, the spacer K may comprise one or more amide, ester, amino, ether, and/or thioether functionalities, 20 and optionally aromatic or mono/polyunsaturated hydrocarbons, polyoxyethylene such as polyethylene glycol, oligo/polyamides such as poly- $\alpha$ -alanine, polyglycine, polylysine, peptides, oligosaccharides, or oligo/polyphosphates. Moreover the spacer may consist of combined units thereof. The length of the spacer may vary, taking into consideration the desired or necessary positioning and spatial orientation of the "active/functional" part of the 25 group in question in relation to the 5- or 6-membered ring. In particularl embodiments, the spacer includes a chemically cleavable group. Examples of such chemically cleavable groups include disulphide groups cleavable under reductive conditions and peptide fragments cleavable by peptidases.

By "target nucleic acid" or "nucleic acid target" is meant a particular nucleic acid 30 sequence of interest. Thus, the "target" can exist in the presence of other nucleic acid molecules or within a larger nucleic acid molecule.

By "solid support" is meant any rigid or semi-rigid material to which a nucleic acid binds or is directly or indirectly attached. The support can be any porous or non-porous water insoluble material, including without limitation, membranes, filters, chips, slides, wafers, fibers, 35 magnetic or nonmagnetic beads, gels, tubing, strips, plates, rods, polymers, particles,

microparticles, capillaries, and the like. The support can have a variety of surface forms, such as wells, trenches, pins, channels and pores.

By an "array" is meant a fixed pattern of at least two different immobilized nucleic acids on a solid support. Desirably, the array includes at least  $10^2$ , more desirably, at least  $10^3$ , 5 and, most desirably, at least  $10^4$  different nucleic acids.

By "antisense nucleic acid" is meant a nucleic acid, regardless of length, that is complementary to a coding strand or mRNA of interest. In some embodiments, the antisense molecule inhibits the expression of only one nucleic acid, and in other embodiments, the antisense molecule inhibits the expression of more than one nucleic acid. Desirably, the 10 antisense nucleic acid decreases the expression or biological activity of a nucleic acid and/or encoded protein by at least 20, 40, 50, 60, 70, 80, 90, 95, or 100%. An antisense molecule can be introduced, e.g., to an individual cell or to whole animals, for example, it may be introduced systemically via the bloodstream. Desirably, a region of the antisense nucleic acid or the entire antisense nucleic acid is at least 70, 80, 90, 95, 98, or 100% complementary to a coding 15 sequence, regulatory region (5' or 3' untranslated region), or an mRNA of interest. Desirably, the region of complementarity includes at least 5, 10, 20, 30, 50, 75, 100, 200, 500, 1000, 2000 or 5000 nucleotides or includes all of the nucleotides in the antisense nucleic acid.

In some embodiments, the antisense molecule is less than 200, 150, 100, 75, 50, or 25 nucleotides in length. In other embodiments, the antisense molecule is less than 50,000; 20 10,000; 5,000; or 2,000 nucleotides in length. In certain embodiments, the antisense molecule is at least 200, 300, 500, 1000, or 5000 nucleotides in length. In some embodiments, the number of nucleotides in the antisense molecule is contained in one of the following ranges: 5-15 nucleotides, 16-20 nucleotides, 21-25 nucleotides, 26-35 nucleotides, 36-45 nucleotides, 46-60 nucleotides, 61-80 nucleotides, 81-100 nucleotides, 101-150 nucleotides, or 151-200 25 nucleotides, inclusive. In addition, the antisense molecule may contain a sequence that is less than a full-length sequence or may contain a full-length sequence.

By "double stranded nucleic acid" is meant a nucleic acid containing a region of two or more nucleotides that are in a double stranded conformation. In various embodiments, the double stranded nucleic acid consists entirely of LNA units or a mixture of LNA units, 30 ribonucleotides, and/or deoxynucleotides. The double stranded nucleic acid may be a single molecule with a region of self-complementarity such that nucleotides in one segment of the molecule base-pair with nucleotides in another segment of the molecule. Alternatively, the double stranded nucleic acid may include two different strands that have a region of complementarity to each other. Desirably, the regions of complementarity are at least 70, 80, 35 90, 95, 98, or 100% complementary. Desirably, the region of the double stranded nucleic acid

that is present in a double stranded conformation includes at least 5, 10, 20, 30, 50, 75, 100, 200, 500, 1000, 2000 or 5000 nucleotides or includes all of the nucleotides in the double stranded nucleic acid. Desirable double stranded nucleic acid molecules have a strand or region that is at least 70, 80, 90, 95, 98, or 100% identical to a coding region or a regulatory sequence (e.g., a transcription factor binding site, a promoter, or a 5' or 3' untranslated region) of a nucleic acid of interest. In some embodiments, the double stranded nucleic acid is less than 200, 150, 100, 75, 50, or 25 nucleotides in length. In other embodiments, the double stranded nucleic acid is less than 50,000; 10,000; 5,000; or 2,000 nucleotides in length. In certain embodiments, the double stranded nucleic acid is at least 200, 300, 500, 1000, or 5000 nucleotides in length. In some embodiments, the number of nucleotides in the double stranded nucleic acid is contained in one of the following ranges: 5-15 nucleotides, 16-20 nucleotides, 21-25 nucleotides, 26-35 nucleotides, 36-45 nucleotides, 46-60 nucleotides, 61-80 nucleotides, 81-100 nucleotides, 101-150 nucleotides, or 151-200 nucleotides, inclusive. In addition, the double stranded nucleic acid may contain a sequence that is less than a full-length sequence or may contain a full-length sequence.

In some embodiments, the double stranded nucleic acid inhibits the expression of only one nucleic acid, and in other embodiments, the double stranded nucleic acid molecule inhibits the expression of more than one nucleic acid. Desirably, the nucleic acid decreases the expression or biological activity of a nucleic acid of interest or a protein encoded by a nucleic acid of interest by at least 20, 40, 50, 60, 70, 80, 90, 95, or 100%. A double stranded nucleic acid can be introduced, e.g., to an individual cell or to whole animals, for example, it may be introduced systemically via the bloodstream.

In various embodiments, the double stranded nucleic acid or antisense molecule includes one or more LNA nucleotides, one or more universal bases, and/or one or more modified nucleotides in which the 2' position in the sugar (e.g., ribose or xylose) contains a halogen (such as fluorine group) or contains an alkoxy group (such as a methoxy group) which increases the half-life of the double stranded nucleic acid or antisense molecule *in vitro* or *in vivo* compared to the corresponding double stranded nucleic acid or antisense molecule in which the corresponding 2' position contains a hydrogen or an hydroxyl group. In yet other embodiments, the double stranded nucleic acid or antisense molecule includes one or more linkages between adjacent nucleotides other than a naturally-occurring phosphodiester linkage. Examples of such linkages include phosphoramidate, phosphorothioate, and phosphorodithioate linkages. Desirably, the double stranded or antisense molecule is purified.

By "purified" is meant separated from other components that naturally accompany it. Typically, a factor is substantially pure when it is at least 50%, by weight, free from proteins,

antibodies, and naturally-occurring organic molecules with which it is naturally associated. Desirably, the factor is at least 75%, more desirably, at least 90%, and most desirably, at least 99%, by weight, pure. A substantially pure factor may be obtained by chemical synthesis, separation of the factor from natural sources, or production of the factor in a recombinant host cell that does not naturally produce the factor. Nucleic acids and proteins may be purified by one skilled in the art using standard techniques such as those described by Ausubel *et al.* (Current Protocols in Molecular Biology, John Wiley & Sons, New York, 2000). The factor is desirably at least 2, 5, or 10 times as pure as the starting material, as measured using polyacrylamide gel electrophoresis, column chromatography, optical density, HPLC analysis, 10 or western analysis (Ausubel *et al.*, *supra*). Desirable methods of purification include immunoprecipitation, column chromatography such as immunoaffinity chromatography, magnetic bead immunoaffinity purification, and panning with a plate-bound antibody.

By "treating, stabilizing, or preventing a disease, disorder, or condition" is meant preventing or delaying an initial or subsequent occurrence of a disease, disorder, or condition; 15 increasing the disease-free survival time between the disappearance of a condition and its reoccurrence; stabilizing or reducing an adverse symptom associated with a condition; or inhibiting or stabilizing the progression of a condition. Desirably, at least 20, 40, 60, 80, 90, or 95% of the treated subjects have a complete remission in which all evidence of the disease disappears. In another desirable embodiment, the length of time a patient survives after being 20 diagnosed with a condition and treated with a nucleic acid of the invention is at least 20, 40, 60, 80, 100, 200, or even 500% greater than (i) the average amount of time an untreated patient survives or (ii) the average amount of time a patient treated with another therapy survives.

By "treating, stabilizing, or preventing cancer" is meant causing a reduction in the size of a tumor, slowing or preventing an increase in the size of a tumor, increasing the disease-free 25 survival time between the disappearance of a tumor and its reappearance, preventing an initial or subsequent occurrence of a tumor, or reducing an adverse symptom associated with a tumor. In one desirable embodiment, the number of cancerous cells surviving the treatment is at least 20, 40, 60, 80, or 100% lower than the initial number of cancerous cells, as measured using any standard assay. Desirably, the decrease in the number of cancerous cells induced by 30 administration of a nucleic acid of the invention (e.g., a nucleic acid with substantial complementarity to a nucleic acid associated with cancer such as an oncogene) is at least 2, 5, 10, 20, or 50-fold greater than the decrease in the number of non-cancerous cells. In yet another desirable embodiment, the number of cancerous cells present after administration of a nucleic acid of the invention is at least 2, 5, 10, 20, or 50-fold lower than the number of 35 cancerous cells present prior to the administration of the compound or after administration of a

buffer control. Desirably, the methods of the present invention result in a decrease of 20, 40, 60, 80, or 100% in the size of a tumor as determined using standard methods. Desirably, at least 20, 40, 60, 80, 90, or 95% of the treated subjects have a complete remission in which all evidence of the cancer disappears. Desirably, the cancer does not reappear or reappears after at least 5, 10, 15, or 20 years.

Exemplary cancers that can be treated, stabilized, or prevented using the above methods include prostate cancers, breast cancers, ovarian cancers, pancreatic cancers, gastric cancers, bladder cancers, salivary gland carcinomas, gastrointestinal cancers, lung cancers, colon cancers, melanomas, brain tumors, leukemias, lymphomas, and carcinomas. Benign tumors may also be treated or prevented using the methods and nucleic acids of the present invention.

By "infection" is meant the invasion of a host animal by a pathogen (e.g., a bacteria, yeast, or virus). For example, the infection may include the excessive growth of a pathogen that is normally present in or on the body of an animal or growth of a pathogen that is not normally present in or on the animal. More generally, an infection can be any situation in which the presence of a pathogen population(s) is damaging to a host. Thus, an animal is "suffering" from an infection when an excessive amount of a pathogen population is present in or on the animal's body, or when the presence of a pathogen population(s) is damaging the cells or other tissue of the animal. In one embodiment, the number of a particular genus or species of pathogen is at least 2, 4, 6, or 8 times the number normally found in the animal.

At bacterial infection may be due to gram positive and/or gram negative bacteria. In desirable embodiments, the bacterial infection is due to one or more of the following bacteria: *Chlamydophila pneumoniae*, *C. psittaci*, *C. abortus*, *Chlamydia trachomatis*, *Simkania negevensis*, *Parachlamydia acanthamoebae*, *Pseudomonas aeruginosa*, *P. alcaligenes*, *P. chlororaphis*, *P. fluorescens*, *P. luteola*, *P. mendocina*, *P. monteilii*, *P. oryzae*, *P. pertucinogena*, *P. pseudocaligenes*, *P. putida*, *P. stutzeri*, *Burkholderia cepacia*, *Aeromonas hydrophilia*, *Escherichia coli*, *Citrobacter freundii*, *Salmonella typhimurium*, *S. typhi*, *S. paratyphi*, *S. enteritidis*, *Shigella dysenteriae*, *S. flexneri*, *S. sonnei*, *Enterobacter cloacae*, *E. aerogenes*, *Klebsiella pneumoniae*, *K. oxytoca*, *Serratia marcescens*, *Francisella tularensis*, *Morganella morganii*, *Proteus mirabilis*, *Proteus vulgaris*, *Providencia alcalifaciens*, *P. rettgeri*, *P. stuartii*, *Acinetobacter calcoaceticus*, *A. haemolyticus*, *Yersinia enterocolitica*, *Y. pestis*, *Y. pseudotuberculosis*, *Y. intermedia*, *Bordetella pertussis*, *B. parapertussis*, *B. bronchiseptica*, *Haemophilus influenzae*, *H. parainfluenzae*, *H. haemolyticus*, *H. parahaemolyticus*, *H. ducreyi*, *Pasteurella multocida*, *P. haemolytica*, *Branhamella catarrhalis*, *Helicobacter pylori*, *Campylobacter fetus*, *C. jejuni*, *C. coli*, *Borrelia burgdorferi*, *V. cholerae*, *V. parahaemolyticus*, *Legionella pneumophila*, *Listeria monocytogenes*, *Neisseria gonorrhea*,



*N. meningitidis*, *Kingella dentrificans*, *K. kingae*, *K. oralis*, *Moraxella catarrhalis*, *M. atlantae*, *M. lacunata*, *M. nonliquefaciens*, *M. osloensis*, *M. phenylpyruvica*, *Gardnerella vaginalis*, *Bacteroides fragilis*, *Bacteroides distasonis*, *Bacteroides 3452A* homology group, *Bacteroides vulgatus*, *B. ovalus*, *B. thetaiotaomicron*, *B. uniformis*, *B. eggerthii*, *B. splanchnicus*,  
 5 *Clostridium difficile*, *Mycobacterium tuberculosis*, *M. avium*, *M. intracellulare*, *M. leprae*, *C. diphtheriae*, *C. ulcerans*, *C. accolens*, *C. afermentans*, *C. amycolatum*, *C. argentorensense*, *C. auris*, *C. bovis*, *C. confusum*, *C. coyleae*, *C. durum*, *C. falsenii*, *C. glucuronolyticum*, *C. imitans*, *C. jeikeium*, *C. kutscheri*, *C. kroppenstedtii*, *C. lipophilum*, *C. macginleyi*, *C. matruchoti*, *C. mucifaciens*, *C. pilosum*, *C. propinquum*, *C. renale*, *C. riegelii*, *C. sanguinis*, *C.*  
 10 *singulare*, *C. striatum*, *C. sundsvallense*, *C. thomssenii*, *C. urealyticum*, *C. xerosis*, *Streptococcus pneumoniae*, *S. agalactiae*, *S. pyogenes*, *Enterococcus avium*, *E. casseliflavus*, *E. cecorum*, *E. dispar*, *E. durans*, *E. faecalis*, *E. faecium*, *E. flavescens*, *E. gallinarum*, *E. hirae*, *E. malodoratus*, *E. mundtii*, *E. pseudoavium*, *E. raffinosus*, *E. solitarius*, *Staphylococcus aureus*, *S. epidermidis*, *S. saprophyticus*, *S. intermedius*, *S. hyicus*, *S. haemolyticus*, *S. hominis*, and/or  
 15 *S. saccharolyticus*. Desirably, a nucleic acid is administered in an amount sufficient to prevent, stabilize, or inhibit the growth of a pathogenic bacteria or to kill the bacteria.

In various embodiments, the viral infection relevant to the methods of the invention is an infection by one or more of the following viruses: West Nile virus (e.g., Samuel, "Host genetic variability and West Nile virus susceptibility," Proc. Natl. Acad. Sci. USA August 21,  
 20 2002; Beasley, Virology 296:17-23, 2002), Hepatitis, picornavirus, polio, HIV, coxsackie, herpes (e.g., zoster, simplex, EBV, or CMV), adenovirus, retrovirus, falvi, pox, rhabdovirus, picorna virus (e.g., coxsackie, entero, hoof and mouth, polio, or rhinovirus), St. Louis encephalitis, Epstein-Barr, myxovirus, JC, coxsackievirus B, togavirus, measles, paramyxovirus, echovirus, bunyavirus, cytomegalovirus, varicella-zoster, mumps, equine encephalitis,  
 25 lymphocytic choriomeningitis, rabies, simian virus 40, polyoma virus, parvovirus, papilloma virus, primate adenovirus, and/or BK.

By "mammal in need of treatment" is meant a mammal in which a disease, disorder, or condition is treated, stabilized, or prevented by the administration of a nucleic acid of the invention.

30 Other aspects and embodiments of the invention are in the detailed description and claims below. Additionally, other nucleic acids and methods described in U.S.S.N. 10/105,639 (Jakobsen *et al.*, "Modified Oligonucleotides and Uses Thereof") or U.S.S.N. 60/410,061 (Ramsing *et al.*, "Populations of Oligonucleotides with Duplex Stabilizing Properties and Uses Thereof") which are hereby incorporated by reference, can be used in the present invention. [  
 35

### Brief Description Of The Drawings

The application file contains drawings executed in color (Figs. 14, 15, 16A, 16B, 17, 18A, 18B, 22, 27, 29, 32, 34, 35A, 35B, 36A, 36B, and 37).

5 Copies of this patent or patent application with color drawings will be provided by the Office upon request and payment of the necessary fee.

Figure 1 shows the structures of selected nucleotide monomers: DNA (T), LNA ( $T^L$ ), pyrene DNA (Py), 2'-OMe-RNA [2'-OMe(T)], abasic LNA ( $Ab^L$ ), phenyl LNA (17a), and pyrenyl LNA (17d).

10 Figure 2 illustrates the chemical structures of Selective Binding Complementary (SBC) nucleotides.

Figure 3 is a table comparing the  $T_m$  differential of various DNA oligomers to LNA oligomers.

Figure 4 shows the base-pairing between modified bases and naturally-occurring 15 nucleotides (EP 1 072 679). These modified bases may be incorporated as part of an LNA, DNA, or RNA unit and used any of the oligomers of the invention.

Figure 5 shows the structure of desirable adenosine analogs (WO 97/12896). These modified bases may be incorporated as part of an LNA, DNA, or RNA unit and used any of the oligomers of the invention.

20 Figure 6 shows the structure of desirable thymine analogs (WO 97/12896). These modified bases may be incorporated as part of an LNA, DNA, or RNA unit and used any of the oligomers of the invention.

Figure 7 shows the structure of desirable guanine analogs (WO 97/12896). These modified bases may be incorporated as part of an LNA, DNA, or RNA unit and used any of the 25 oligomers of the invention.

Figure 8 shows the structure of desirable cytosine analogs (WO 97/12896). These modified bases may be incorporated as part of an LNA, DNA, or RNA unit and used in any of the oligomers of the invention.

Figure 9 is a schematic illustration of the use of an exemplary synthesis for LNA-30 furanoPyr-SBC-C.

Figure 10 is a schematic illustration of the use of nucleic acid probes with secondary structure (e.g., hairpins formed due to flanking universal base). The nucleic acid probe contains both a fluorophore and a quencher such that the fluorescence intensity of the fluorophore is decreased by a nearby quencher. When a test nucleic acid is bound to the capture probe, the 35 level of fluorescence intensity increases because of the increased distance between the

fluorophore and quencher. This method allows a test nucleic acid sample to be characterized and quantitated without being labeled directly. The amount of capture probe in a particular spot in the array can also be measured by heating the array to reduce secondary structure (e.g., eliminate hairpins) and measuring the fluorescence.

5 Figure 11 is a schematic illustration of the synthesis of some compounds of the invention.

Figure 12A and 12 B show the sensitivity of 50-mer LNA capture probes compared to 50-mer DNA capture probes. SW15-specific 50-mer DNA oligonucleotides (green bars) and 50-mer capture probes with an LNA nucleotide incorporated at every third nucleotide position 10 (red bars) were printed at the oligo concentration indicated below. The slides were hybridized at 65°C in 3x SSC (Fig. 12A) and at 70°C in 3xSSC (Fig. 12B).

Figures 13A and 13B show the specificity of 40-mer LNA capture probes (red bars) compared to DNA capture probes (green bars). The hybridizations were carried out at 65°C in 3XSSC. Bars 1 and 7 represent perfectly matched duplexes, bars 2 and 8, 3, and 9, 4 and 10, 5 15 and 11, and 6 and 12 represent duplexes with 1, 2, 3, 4, 5 mismatches, respectively. The *in vitro* RNA used was SW15 in Fig 13A and TH14 in Fig. 13B.

Fig. 14 shows the detection principle for alternative exon skipping in the *C. elegans let-2* gene using LNA oligonucleotide capture probes and comparative expression profiling.

Figure 15 shows the detection of alternative splicing of *C. elegans* Let-2 exon 9 and 10 20 using LNA-modified capture probes.

Figures 16A and 16B show the comparison of DNA and LNA-modified oligonucleotide capture probes in the specific capture of the *C. elegans* gene26/T01D3.3 exon 4.

Figure 17 illustrates the LNA exon-exon junction (merged) probe concept.

Figures 18A and 18B show the capture probe specificity for the *C. elegans* 25 gene26/T01D3.3 exon 4 (Fig. 18A) and exon 5 (Fig. 18B) as validated by short complementary target oligonucleotides.

Figure 19 shows the construction of the recombinant splice variants in the *in vitro* transcription vector. The small bars show the location of the hybridization for the oligonucleotide capture probes used in this example. The sequences of the capture probes are 30 described herein.

Figures 20A and 20B or bar graphs show the detection of splice variant #1 and #2, respectively using LNA-modified capture probes in a comparative hybridization.

Fig. 21 shows the sensitivity of 50-mer LNA capture probes compared to 50-mer DNA capture probes. SW15-specific 50-mer DNA oligonucleotides and 50-mer capture probes with

an LNA nucleotide incorporated at every second (LNA2) or third (LNA3) nucleotide position. The slides were hybridized at 65°C in 3×SSC.

Figure 22 is a bar graph of the signal intensities of a patient DNA sample hybridized to an array of the invention. The names of the probes in Figs. 22 and 25 match, although the 5 numbers used in Fig. 22 are abbreviated, e.g., probe No. 10580 Menkes.14 50NH2C6-2.LNA in Fig. 25 corresponds to the second probe counted from the left "14.2" LNA in the lower graph of Fig. 22.

Figure 23 is a graph comparing the spot intensity for probes of the invention with different LNA substitution patterns.

10 Figure 24 is a bar graph of the spot intensity for LNA probes for different exons.

Figure 25 is a table of CGH capture probe sequences.

Figure 26 is a flow chart of the steps of oligo design software of the invention. The OligoDesign software features LNA modified oligonucleotide secondary structure prediction, LNA spiked oligo melting temperature prediction, genome wide cross hybridization prediction, 15 secondary structure prediction of the target, and recognition and filtering of the target in the genome. These features are determined for each possible probe of the query gene and presented to an artificial neural network. The probes are then ranked according to the neural network prediction and the top scoring probes are returned.

Figure 27 is a schematic illustration of oligo design software of the invention.

20 Figures 28A-28E contain computer code for exemplary software of the invention.

The oligod program takes a gene sequence as input and returns sequences for LNA spiked oligonucleotides (Fig. 28A). The dyp program is used by oligod to predict the secondary structure and self annealing properties of the oligonucleotides (Fig. 28B). The expression\_array\_param file contains parameters used by the oligod program (Fig. 28C). Fig. 25 28D contains code for a  $T_m$  prediction program, and Fig. 28E contains code for a  $T_m$  thermodynamic model.

Figure 29 illustrates photo-activated immobilization of nucleic acids of the invention, which enables polarized coupling of anthraquinone (AQ)-lined LNA oligonucleotides onto the polymer surface. No pretreatment of the slide is needed. A covalent bond is formed between 30 the oligonucleotide and the polymer using a UV source, e.g. Stratalinker.

Figure 30 illustrates an injection-molded polymer slide. Finger indents ease slide handling. The slide has a well-defined printing and hybridization window, frosted surface for identification and orientation, and space for barcodes.

Figure 31 illustrates spot quality on different slides that can be used to immobilize 35 nucleic acids of the invention. The hydrophobic slide surface ensures that extremely

homogenous spots are generated when hydrophilic spotting solution is applied to the surface. A high spot quality is obtained on the EURAY™ polymer slide compared to a glass slide when using a spot-to-spot distance of 150 :M. The high-quality arrays simplify downstream image analyses.

5 Figure 32 is a schematic illustration of a method of the invention.

Figure 33 is a table of exemplary target nucleic acids (Holstege et al. (1998)( Cell 95, 717-728, and Causton et al. (2001) Mol. Biol. Cell 12, 323-337).

Figure 34 is a graph of Cy5 intensity. Yeast actin 1-specific 50-mer capture probes were synthesized as DNA and DNA/LNA mixmer oligonucleotides. LNA mixmer capture  
10 probes contain an LNA at every 4<sup>th</sup>, 5<sup>th</sup>, and 6<sup>th</sup> nucleotide position (LNA\_4, LNA\_5, LNA\_6). On-chip melting profiles demonstrate a 8-10 °C increase in T<sub>m</sub> obtained with LNA capture probes.

Figure 35A illustrates the heat-shock response in yeast. The array was hybridized with Cy3-labeled standard and Cy5-labelled heat-shock yeast cDNA. Figure 35B also illustrates the  
15 heat-shock response in yeast. The microarray data were normalized using yeast actin 1. The *ssa4* gene encoding heat shock protein HSP70 is up-regulated over 2-fold. Expression of the *gual* gene is down-regulated.

Figure 36A compares expression of wild-type and *ssa4* mutant yeast. The array was hybridized with Cy3-labeled wild-type and Cy5-labelled *ssa4* mutant yeast cDNA. Figure 36B  
20 also compares wild-type and *ssa4* yeast. The hybridization data were normalized using yeast actin 1. *ssa4* is detected in the wild-type yeast strain, but not in the *ssa4* knock-out strain.

Figure 37 illustrates mRNA splicing.

Figure 38 is a picture showing gel electrophoresis of fragmented cDNA from the yeast wild-type strain. The molecular marker (lane 1 and 9) is from Life technologies, USA. Lanes  
25 2-8 represents the UDG-fragmented cDNA 1-7 according to the different dUTP/dTTP ratios in Table 18.

Figure 39 is a graph of the log ratios of the normalized fluorescence intensities from the wild-type yeast strain (signal) and those from the  $\Delta$ *ssa4* yeast strain (noise) as a function of capture probe position in the 3' region of the SSA4 mRNA.

30 Figure 40 is a schematic illustration of mRNA splicing.

Figure 41 is a schematic illustration of alternative mRNA splicing.

Figure 42 is a schematic illustration of probes of the invention.

Figure 43 is a schematic illustration of probes of the invention.

Figure 44 illustrates an exemplary computer for use in the methods of the invention.

### Detailed Description

#### *Detection and Analysis of mRNA Splice Variants*

Alternative splicing is the process by which different mature messenger RNAs are produced from the same pre-mRNA. Because the mRNA composition of a given cell determines the proteins present in a cell, this process is an important aspect of a cell's gene expression profile. Current investigations of transcriptomes (i.e., the total complexity of RNA transcripts produced by an organism) indicate that at least 50-60 % of the genes of complex eukaryotes produce more than one splice variant. The present invention provides a novel method for detecting and quantifying the levels of splice variants in complex mRNA pools using LNA discriminating probes and high-throughput LNA oligonucleotide microarray technology. The detection concept which uses internal LNA exon probes and/or splice-variant specific exon-exon junction or exon-intron or intron-exon (so-called merged) probes is depicted in Fig. 17.

Internal, exon-specific (or intron-specific) LNA oligonucleotide probes are designed and used to detect the relative levels of a given exon (or intron) in complex mRNA pools using oligonucleotide microarray technology or similar techniques. Exon-exon LNA junction probes are designed for multiple or all possible exon-exon combinations or exon-intron combinations. The LNA discriminating probes are highly specific and superior compared to DNA oligos due to the higher  $\Delta T_m$  of LNA probes. These probes can be used to determine the sequential order of each sub-element (i.e., exon structure or exon-intron structure) in a given alternatively spliced mRNA isoform, thus giving the exact composition of the mRNA. Subsequently, the ratios of each splice variant can be quantified using the combined readouts from both internal and merged LNA probes and control probes. The invention is applicable both in single fluor (single channel) or comparative two-fluor (two channel) microarray hybridizations.

Several "artificial," alternatively spliced mRNA molecules may be constructed in an *in vitro* transcription vector for the production of clean IVT RNA. Both internal and junction-specific LNA oligonucleotide capture probes are designed, synthesized, and spotted onto, e.g., Exiqon's polymer microarray platform. The resulting splice-specific microarray are used to validate the LNA discriminating probe concept by spiking the *in vitro* RNAs individually as well as in different ratios into a complex RNA background for fluorochrome-labelling and array hybridization.

The internal and merged probes of the invention can also be used in any standard method for the analysis of mRNA splice variants (see, for example, Yeakley *et al.*, Nature Biotechnology 20:353-358, 2002; Clark *et al.*, Science 296:907-910, 2002; Mutch *et al.*, Genome Biology 2(12):preprint00009.1-0009.31, 2001).

### *Exemplary Applications of Internal and/or Merged Probes*

The internal and/or merged probes of the invention can also be used for gene expression profiling of alternative splice variants, oligonucleotide expression microarrays, real-time PCR, 5 and profiling of alternatively spliced mRNAs using microtiterplate assays or fiber-optic arrays.

Detection and characterization of alternative splicing is particularly useful for the study and treatment of human disease (exonhit website, "Inaugural Splicing 2002 Concludes: Alternative Splicing May Make All the Difference in Discovering the Origin of Disease). In particular, RNA splicing is now widely recognized as a means to generate protein diversity. 10 Alternative splicing is a key mechanism for regulating gene expression, and any mutation or defect in its regulation can impact considerably cell functions. Therefore, it is likely to be an important source of novel gene and protein targets implicated in human pathology. Industry has long recognized the need for innovative discovery technologies that focuses on the origin of disease for the development of novel diagnostics and therapeutics.

15 In particular, there are many examples of human pathologies caused by alterations in normal patterns of alternative RNA splicing. Because a large number of human genes undergo alternative splicing, the protein isoforms that result from this process represent a major source of targets for commercial development of therapies and diagnostics. In particular, splicing processes play a significant role in the onset and development of cardiovascular, muscular, 20 CNS diseases, and cancer. Early evidence indicates the origin of many diseases can be identified by examining alternative splicing – which leads to the point of intervention for discovering future generations of drugs. The study of splicing enables the discovery of new mechanisms underlying disease progression.

### *Comparative Genomic Hybridization*

25 Comparative Genomic Hybridization (CGH) is a powerful technology for detection of unbalanced chromosome rearrangements and holds much promise for screening and identification of interstitial submicroscopic rearrangements that otherwise cannot be detected using classical cytogenetic or FISH technologies. The adaptation of CGH onto an oligo microarray platform allows detection of small single exon deletion/duplications on a genome 30 wide scale. There is a strong need for developing microarrays that can detect, e.g., single exon aberrations. This detection can be achieved by employing LNA mixmer oligos as capture probes for individual exons in selected genes.

A model system for these methods is the Menkes loci. Menkes disease is a lethal-X linked recessive disorder associated with copper metabolism disturbance leading to death in early 35 childhood. The Menkes locus has been mapped to Xq13. The gene spans about 150 kb

genomic region, contains 23 exons, and encodes a 8.5 kb gene transcript. The gene for Menkes disease (now designated as *ATP7A*) encodes a 1500 amino acid membrane-bound Cu-binding P-type ATPase (*ATP7A*). The 8.5 kb transcript is expressed in all tissues from normal individuals (though only trace amounts are present in liver), but is diminished or absent in 5 Menkes disease patients. Several different kinds of mutations, like chromosome aberrations, point mutations and partial gene deletions affecting *ATP7A* have been identified in MD patients. 50-mer capture probes with LNA spiked in every second, third, and fourth position have been designed for every exon (23 exons) representing *ATP7A*, using the OligoDesign software tool, described herein. The C6-amino-linked capture probes were spotted onto 10 Immobilizer slides and hybridized with patient samples with Cy3 fluorescent dye and a known reference genomic sample with Cy5. After mixing equal amounts of the labelled DNA, the probe is hybridized to array. The ratio of Cy5 signal to Cy3 for each clone indicates differences in chromosome/DNA material. For example, the Cy5 signal is higher than Cy3 if the patient genome has a deletion, and is lower if there is duplication. In regions that are 15 unchanged, the Cy5: Cy3 ratio is 1:1. These methods can be used to analyze a number of well-characterized Menkes patients with a range of partial deletions of *ATP7A*.

LNA oligonucleotide-based CGH makes it possible to assess a large number of chromosomal aberrations that are being screen for in the cytogenetic clinic. In contrast, standard FISH analysis typically only detects large chromosomal rearrangements. In desirable 20 embodiments, an array that contains a series of overlapping probes is used to detect a chromosomal deletion in a nucleic acid sample, such as a patient sample.

### ***Clinical diagnostics***

Clinical diagnosis is a key element in healthcare management and point-of-care. A large number of analyses in the hospitals are based on the use of robust, cost efficient, sensitive and 25 highly specific diagnostic tests. Thus, the diagnosis of various diseases is performed with a high selectivity and reliability, resulting in confirmation of medical diagnosis, choice of therapy and follow-up treatment as well as prevention. In addition to its importance in the quality of healthcare provided to patients, clinical diagnosis also contributes to the control of healthcare costs. The field of clinical diagnostics involves analyzing biological fluid samples (blood, urine, 30 etc.) or biopsies collected from patients in order to establish the diagnosis of diseases, whether of infectious, metabolic, endocrine or cancerous origin. Medical analysis of infectious diseases involves testing and identifying the micro-organisms causing the infection e.g. testing for and identifying a micro-organism in blood and determining its susceptibility to antibiotics or detecting an antigen-antibody reaction produced as a response to an attack by a micro-organism 35 in the human body, e.g. testing for antibodies for the diagnosis of hepatitis. The accurate



diagnosis of metabolic and endocrine diseases and cancers, resulting in a disease phenotype with a bodily imbalance, involves the measurement of diagnostic substances or elements present in the biological fluids or biopsies. These substances are examined and results are interpreted with reference to known normal values.

##### 5 *Use of diagnostic kits in microbiological control*

The pharmaceutical, cosmetics and agri-food industries are being confronted with increasingly strict quality standards. Thus, the purpose of industrial microbiological control testing is to detect and measure the presence of potentially pathogenic microbial contaminants throughout the manufacturing process from raw materials to the finished products, as well as in the  
10 production environment. The obtained results are subsequently compared to the current regulatory guidelines and industry standards.

##### *Application of molecular biological techniques to in vitro diagnostics*

Recently, several different molecular biological techniques have been used successfully in accurate quantification of RNA levels in clinical diagnosis as well as in microbiological control.  
15 The applications are wide-ranging and include methods for quantification of the regulation and expression of drug resistance markers in tumour cells, monitoring of the responses to chemotherapy, measuring the biodistribution and transcription of gene-encoded therapeutics, molecular assessment of the tumor stage in a given cancer, detecting circulating tumor cells in cancer patients and detection of bacterial and viral pathogens. The reverse transcription  
20 polymerase chain reaction (RT-PCR) is the most sensitive method for the detection of mRNA, including low abundant mRNAs, often obtained from limited tissue samples in clinical diagnostics. The application of fluorescence techniques to RT-PCR combined with suitable instrumentation has led to development of quantitative RT-PCR methods, combining amplification, detection and quantification in a closed system avoid from contamination and  
25 with minimized hands-on time. The two most commonly used quantitative RT-PCR techniques are the Taqman RT-PCR assay (ABI, Foster City, USA) and the Lightcycler assay (Roche, USA). A third method applied to detection and quantification of RNA levels is real-time nucleic acid sequence based amplification (NASBA) combined with molecular beacon detection molecules. NASBA is a single-step isothermal RNA-specific amplification method  
30 that amplifies mRNA in a double stranded DNA environment, and this method has recently proven useful in the detection of various mRNAs and in the detection of both viral and bacterial RNA in clinical samples. Finally, the recent explosion in microarray technology holds the promise of using microarrays in clinical diagnostics. For example van't Veer et al. (Nature 2002: 415, 31) describe the successful use of microarrays in obtaining digital mRNA signatures

from breast tumors and the use of these signatures in the precise prediction of the clinical outcome of breast cancer in patients.

The success of exploiting molecular biological techniques in diagnostics and diagnostic kits depends on continuous optimization of the technologies and the development of new robust and cost-effective technology platforms for producing accurate, reproducible and valid clinical data. Locked nucleic acid (LNA) oligonucleotides constitute a novel class of bicyclic RNA analogs having an exceptionally high affinity and specificity toward their complementary DNA and RNA target molecules. Besides increased thermal stability, LNA-containing oligonucleotides show significantly increased mismatch discrimination, and allow full control of the melting temperature across microarray hybridizations. The LNA chemistry is completely compatible with conventional DNA phosphoramidite chemistry and thus LNA substituted oligonucleotides can be designed to optimize performance. LNA oligonucleotides would be well-suited for large-scale clinical studies providing highly accurate genotyping by direct competitive hybridization of two allele-specific LNA probes to e.g. microarrays of immobilized patient amplicons. In addition, the use of LNA substituted oligonucleotides would increase both sensitivity and specificity in detection and quantification of mRNA levels in clinical samples, either by quantitative RT-PCR, quantitative NASBA or oligonucleotide microarrays, compared with DNA probes. Application of LNA oligonucleotides into diagnostic kits would thus significantly enhance their performance. Finally, the use of LNA substituted oligonucleotides would increase the sensitivity and specificity in the detection of alternatively spliced mRNA isoforms and non-coding RNAs either by homogeneous assays (Taqman assay, Lightcycler assay, NASBA) or by oligonucleotide microarrays in a massive parallel analysis setup.

#### *Optimized Nucleic Acids of the Invention*

Decreasing the variation in melting temperatures ( $T_m$ ) of a population of nucleic acids allows the nucleic acids to hybridize to target molecules under similar binding conditions, thereby simplifying the simultaneous hybridization of multiple nucleic acids. Similar melting temperatures also allow the same hybridization conditions to be used for multiple experiments, which is particularly useful for assays involving hybridization to nucleic acids of varying "AT" content. For example, current methods often require less stringent conditions for hybridization of nucleic acids with high "AT" content compared to nucleic acids with low "AT" content. Due to this variation in hybridization stringency, current methods may require significant trial and error to optimize the hybridization conditions for each experiment.

To overcome limitations in current nucleic acid hybridization and/or amplification techniques, we have developed populations of nucleic acid probes or primers with minimal variation in melting temperature (U.S.S.N. 60/410,061). For example, the unique properties of

LNA nucleotide analogs increase their binding affinity for DNA and RNA. The stability of duplexes can generally be ranked as follows: DNA:DNA < DNA:RNA < RNA:RNA ≤ LNA:DNA < LNA:RNA < LNA:LNA. The DNA:DNA duplex is thus the least stable and the LNA:LNA duplex the most stable. The affinity of the LNA nucleotides A and T corresponds approximately to the affinity of DNA G and C to their complementary bases. General substitution of one or more A and T nucleotides with LNA A and LNA T in DNA oligonucleotides is therefore a simple way of equalizing differences in  $T_m$ . Furthermore, the mean melting temperature is increased significantly, which is often important for shorter oligonucleotides. For example, predictions of melting temperature of all possible 9-mer oligonucleotides have shown that the mean temperature increases from 39.7 °C to 59.3 °C by substituting all DNA A and T nucleotides with LNA A and T nucleotides. The variance in  $T_m$  of all 9-mers furthermore decreases from 59.6 °C for DNA oligonucleotides to only 4.7 °C for the LNA substituted oligonucleotides. The estimations are based on the latest LNA  $T_m$  prediction algorithms such as those disclosed herein, which have a variance of 6-7 °C.

15 If desired, the capture efficiency of one or more nucleic acids can be increased by including any of the high affinity nucleotides (e.g., LNA units) described herein within the nucleic acids. The examples herein also provide algorithms for optimizing the substitution patterns of the nucleic acids to minimize self-complementarity that may otherwise inhibit the binding of the nucleic acids to target molecules.

20 For various applications of the nucleic acids and arrays of the invention, LNA A and LNA T substitutions are made to equalize the melting temperatures of the nucleic acids. In other embodiments, LNA A and LNA C substitutions are made to minimize self-complementarity and to increase specificity. LNA C and LNA T substitutions also minimize self-complementarity. Additionally, oligonucleotides containing LNA C and LNA T are 25 desirable because these modified nucleotides are easy to synthesis and are especially useful for applications such as antisense technology in which minimizing cost is especially desirable.

The following non-limiting examples are illustrative of the invention. All documents mentioned herein are incorporated herein by reference in their entirety. In the following Examples, compound reference numbers designate the compound as shown in Scheme 1 and 2 30 herein.

#### Example 1: The Use of LNA-modified Oligonucleotides in Microarrays Provide Significantly Improved Sensitivity and Specificity in Expression Profiling

This example demonstrates the advantages of using LNA oligonucleotide microarrays in 35 gene expression profiling experiments. Capture probes for the *Saccharomyces cerevisiae* genes

*SWI5* (YDR146C) and *THI4* (YGR144W) were designed as 50-mer standard DNA and different LNA/DNA "mixmer" oligonucleotides (i.e., oligonucleotides containing both LNA and DNA nucleotides) respectively, for comparison (Table 2). In addition, 40-mer oligonucleotides were designed as truncated versions of the 50-mer capture probes (Table 2).

5 The specificity of the LNA oligoarrays was addressed by introducing 1-5 consecutive mismatches positioned in the middle of 40-mer LNA/DNA mixmer capture probes with LNA in every fourth position. To assess the sensitivity of DNA versus LNA capture probes in complex hybridization mixtures, *in vitro* synthesized yeast RNA for either *SWI5* or *THI4* was spiked into *Caenorhabditis elegans* total RNA for cDNA target synthesis. These experiments are described  
10 further below.

#### *Cultivation of Caenorhabditis elegans worms*

Mixed stage *C. elegans* cultures were grown according to standard methods. Samples were harvested by centrifugation at 3,000×g, suspended in RNA Later storage buffer (Ambion, USA), and immediately frozen in liquid nitrogen.

#### 15 *RNA extraction*

RNA was extracted from the worm samples using the FastRNA® Kit, GREEN (Q-BIO) essentially according to the suppliers' instructions.

#### *In vitro RNA synthesis*

Amplification of the yeast genes was performed using standard PCR with yeast genomic  
20 DNA as the template. In the first step, a forward primer containing a restriction enzyme site and a reverse primer containing a universal linker sequence were used. In the second PCR reaction, the reverse primer was exchanged with a nested primer containing a poly-T20 tail and a restriction enzyme site. The DNA fragments were ligated into the pTRIamp18 vector (Ambion, USA) using the Quick Ligation Kit (New England Biolabs, USA) according to the  
25 supplier's instructions and transformed into *E. coli* DH-5α by standard methods. The PCR clones were sequenced using M13 forward and M13 reverse primers on an ABI 377 (Applied Biosystems, USA). Synthesis of *in vitro* RNA was carried out using the MEGAscript™ T7 Kit (Ambion, USA) according to the manufacturer's instructions.

#### *Design and synthesis of the LNA capture probes*

30 To design the capture probes, regions with unique mRNA sequence of the selected target genes were identified. The optimal 50-mer oligonucleotide sequences with respect to  $T_m$ , self-complementarity, and secondary structure were selected. LNA modifications were incorporated to increase affinity and specificity.

#### *Printing of the LNA microarrays*

The microarrays were printed on Immobilizer™ MicroArray Slides (Exiqon, Denmark) using the Biochip One Arrayer from Packard Biochip technologies (Packard, USA). The arrays were printed with a spot volume of 2 x 300 pl of a 10 µM capture probe solution. Four replicas of the capture probes were printed on each slide.

#### 5 Synthesis of fluorochrome labelled first strand cDNA from total RNA

Ten ng of *S. cerevisiae* *in vitro* synthesized RNA (either *SWI5* or *THI4*) was combined with 10 µg of *C. elegans* total RNA and 5 µg oligo dT primer (T20VN) in an RNase free, pre-siliconized 1.5 mL tube, and the final volume was adjusted with DEPC-water to 8 µL. The reaction mixture was heated at +70°C for 10 minutes, quenched on ice for 5 minutes, and spun 10 for 20 seconds, followed by addition of 1 µL SUPERase-In™ (20U/µL, RNase inhibitor, Ambion, USA), 4 µL 5×RTase buffer (Invitrogen, USA), 2 µL 0.1 M DTT (Invitrogen, USA), 1 µL dNTP (20mM dATP, dGTP, dTTP; 0.4 mM dCTP in DEPC-water, Amersham Pharmacia Biotech, USA), and 3 µL Cy3™-dCTP or Cy5™-dCTP (Amersham Pharmacia Biotech, USA). First strand cDNA synthesis was carried out by adding 1 µL of Superscript™ II (Invitrogen, 15 200 U/mL), mixing, and incubating the reaction mixture for one hour at 42°C. An additional 1 µL of Superscript™ II was added, and the cDNA synthesis reaction mixture was incubated for an additional one hour at 42°C; the reaction was stopped by heating at 70°C for 5 minutes, and quenching on ice for 2 minutes. The RNA was hydrolyzed by adding 3 µL of 0.5 M NaOH, and incubating at 70°C for 15 minutes. The samples were neutralized by adding 3 µL of 0.5 M 20 HCl, and purified by adding 450 µL 1×TE buffer, pH 7.5 to the neutralized sample and transferring the samples onto a Microcon-30 concentrator. The samples were centrifuged at 14000×g in a microcentrifuge for ~8 minutes, the flow-through was discarded and the washing step was repeated twice by refilling the filter with 450 µl 1×TE buffer and by spinning for ~12 minutes. Centrifugation was continued until the volume was reduced to 5 µL, and finally the 25 labelled cDNA probe was eluted by inverting the Microcon-30 tube and spinning at 1000×g for 3 minutes.

#### Hybridization with fluorochrome-labelled cDNA

The arrays were hybridized overnight using the following protocol. The Cy3™ or Cy5™-labelled cDNA samples were combined in one tube followed by addition of 3 µL 30 20×SSC (3×SSC final), 0.5 µL 1 M HEPES, pH 7.0 (25 mM final), 25 µg yeast tRNA (1.25 µg/µL final), 10 µg PolyA blocker (0.5 µg/µL final), 0.6 µL 10% SDS (0.3% final), and DEPC-treated water to 20 µL final volume. The labelled cDNA target sample was filtered in a Millipore 0.22 micron spin column according to the manufacturer's instructions (Millipore,

USA), and the probe was denatured by incubating the reaction at 100°C for 2 minutes. The sample was cooled at 20-25°C for 5 minutes by spinning at maximum speed in a microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was carefully placed on top of the microarray spotted on Immobilizer™ MicroArray Slide, and the hybridization mixture was applied to the array from the side. An aliquot of 30 µL of 3×SSC was added to both ends of the hybridization chamber, and the Immobilizer™ MicroArray Slide was placed in the hybridization chamber. The chamber was sealed watertight and incubated at 65°C for 16-18 hours submerged in a water bath. After hybridization, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The LifterSlip coverslip was washed off in 2 x SSC, pH 7.0 containing 0.1% SDS at room temperature for one minute, followed by washing of the microarrays subsequently in 1.0 x SSC, pH 7.0 at room temperature for one minute, and then in 0.2 x SSC, pH 7.0 at room temperature for one minute. Finally, the slides were washed for 5 seconds in 0.05 x SSC, pH 7.0. The slides were then dried by centrifugation in a swinging bucket rotor at approximately 600 rpm for 5 minutes.

#### 15 Data analysis

Following washing and drying, the slides were scanned using a ScanArray 4000XL scanner (Perkin-Elmer Life Sciences, USA), and the array data were processed using the GenePix™ Pro 4.0 software package (Axon, USA).

#### Results

20 Incorporation of LNA nucleotides at every third nucleotide position in standard 50-mer expression array oligonucleotide capture probes resulted in a 3-fold increase in fluorescence intensity levels, when hybridized under standard stringency conditions (Figs. 12A and 12B). When the hybridization temperature is increased from 65 °C to 70 °C, the capture of the *SWI5* spike mRNA by LNA 50-mer oligos is increased by 8-fold relative to the DNA controls. Thus, it can clearly be concluded that oligonucleotides containing LNA units are more sensitive in expression profiling compared to oligonucleotides containing only DNA units. The specificity of 40-mer LNA/DNA mixmer capture probes in the discrimination of highly homologous target sequences was addressed by introducing 1-5 consecutive mismatches in the middle of *SWI5* and *THI4* capture oligos together with the corresponding DNA controls. As demonstrated in Figs. 13A and 13B, the LNA-spiked (LNA modification at every fourth nucleotide position) 40-mer triple mismatch oligos showed a 3-fold signal intensity decrease relative to the perfectly matched duplexes, whereas the corresponding 40-mer standard DNA capture probes did not form duplexes under standard hybridization stringency. Further, the 40-mer perfect match LNA capture probes showed a 5-fold to 14-fold increase in the intensity levels compared to DNA

oligonucleotides under standard hybridization conditions. Capture probes of other lengths and/or with other LNA substitution patterns can be used similarly.

Table 2. DNA and LNA-modified *SWI5* (YDR146C) and *THI4* (YGR144W) oligonucleotide capture probes. LNA modifications are depicted by uppercase letters in the sequence, mt 5 denotes the number of mismatches (bolded) in the center of the oligonucleotide with respect to its target cDNA (mRNA), and "mC" denotes LNA methyl cytosine.

Oligo Name	Sequence
YDR146C-40	acggggattatggttcgccaatgaaaactaatcaaaggt
YDR146C-40_mt1	acggggattatggttcgcctatgaaaactaatcaaaggt
YDR146C-40_mt2	acggggattatggttcg <b>gt</b> atgaaaactaatcaaaggt
YDR146C-40_mt3	acggggattatggttcg <b>gg</b> atgaaaactaatcaaaggt
YDR146C-40_mt4	acggggattatggttcg <b>gg</b> ttgaaaactaatcaaaggt
YDR146C-40_mt5	acggggattatggttcg <b>gg</b> ttagaaaactaatcaaaggt
YDR146C-40_LNA4	acGgggAttaTggtTtcgmCcaaTgaaAactAatcAaagGt
YDR146C-40_LNA4_mt1	acGgggAttaTggtTtcgmCctaTgaaAactAatcAaagGt
YDR146C-40_LNA4_mt2	acGgggAttaTggtTtcgmC <b>gt</b> aTgaaAactAatcAaagGt
YDR146C-40_LNA4_mt3	acGgggAttaTggtTtcg <b>G</b> taTgaaAactAatcAaagGt
YDR146C-40_LNA4_mt4	acGgggAttaTggtTtcg <b>G</b> ttTgaaAactAatcAaagGt
YDR146C-40_LNA4_mt5	acGgggAttaTggtTtcg <b>G</b> ttAgaaAactAatcAaagGt
YDR146C-50	tgggaatggaacggggattatggttcgccaatgaaaactaatcaaaggt
YDR146C-50_mt1	tgggaatggaacggggattatggtatcgccaatgaaaactaatcaaaggt
YDR146C-50_mt2	tgggaatggaacggggattatggttaacgccaatgaaaactaatcaaaggt
YDR146C-50_mt3	tgggaatggaacggggattatggtta <b>ag</b> ccaatgaaaactaatcaaaggt
YDR146C-50_mt4	tgggaatggaacggggattatgga <b>ag</b> gccaatgaaaactaatcaaaggt
YDR146C-50_mt5	tgggaatggaacggggattatgga <b>ag</b> gcccaatgaaaactaatcaaaggt
YDR146C-50_LNA2	TgGgAaTgGaAcGgGgAtTaTgGtTtmCgmCcAaTgAaAamCtAaTcaAagGt
YDR146C-50_LNA3	TggGaaTggAacGggGatTatGgtTtcGccAatGaaAacTaaTcaAagGt
YDR146C-50_LNA4	TgggAatgGaacGgggAttaTggtTtcgmCcaaTgaaAactAatcAaagGt
YDR146C-50_LNA5	TgggaAtggaAcgggGattaTggtTtcgccAatgaAaactAatcaAaggt
YDR146C-50_LNA6	TgggaaTggaacGgggatTatggtTtcgccAatgaaAactaaTcaaagGt
YDR146C-50_LNA3_mt1	TggGaaTggAacGggGatTatGgtAacGccAatGaaAacTaaTcaAagGt
YDR146C-50_LNA3_mt2	TggGaaTggAacGggGatTatGgt <b>A</b> acGccAatGaaAacTaaTcaAagGt
YDR146C-50_LNA3_mt3	TggGaaTggAacGggGatTatGgt <b>A</b> agGccAatGaaAacTaaTcaAagGt
YDR146C-50_LNA3_mt4	TggGaaTggAacGggGatTatGga <b>A</b> agGccAatGaaAacTaaTcaAagGt
YDR146C-50_LNA3_mt5	TggGaaTggAacGggGatTatGga <b>A</b> agmCccAatGaaAacTaaTcaAagGt
YGR144W-40	ttgctgaactggatggattaaaccgtatgggtccaacttt
YGR144W-40_mt1	ttgctgaactggatggattaaaccgtatgggtccaacttt
YGR144W-40_mt2	<b>ttgctgaactggatggatataaccgtatgggtccaacttt</b>
YGR144W-40_mt3	<b>ttgctgaactggatggatattaccgtatgggtccaacttt</b>

YGR144W-40_mt4	ttgctgaactggatggatatttccgtatgggtccaacttt
YGR144W-40_mt5	ttgctgaactggatggatatttgcgtatgggtccaacttt
YGR144W-40_LNA4	ttGctgAactGgatGgatTaaamCcgtAtggGtccAactTt
YGR144W-40_LNA4_mt1	ttGctgAactGgatGgatTtaamCcgtAtggGtccAactTt
YGR144W-40_LNA4_mt2	ttGctgAactGgatGgatAtaamCcgtAtggGtccAactTt
YGR144W-40_LNA4_mt3	ttGctgAactGgatGgatAttamCcgtAtggGtccAactTt
YGR144W-40_LNA4_mt4	ttGctgAactGgatGgatAtttmCcgtAtggGtccAactTt
YGR144W-40_LNA4_mt5	ttGctgAactGgatGgatAtttGcgtAtggGtccAactTt
YGR144W-50	gggatggaagttgctgaactggatggattaaccgtatgggtccaacttt
YGR144W-50_mt1	gggatggaagttgctgaactggatcgattaaccgtatgggtccaacttt
YGR144W-50_mt2	gggatggaagttgctgaactggatccattaaaccgtatgggtccaacttt
YGR144W-50_mt3	gggatggaagttgctgaactggaaccattaaaccgtatgggtccaacttt
YGR144W-50_mt4	gggatggaagttgctgaactggaaccttaaaaccgtatgggtccaacttt
YGR144W-50_mt5	gggatggaagttgctgaactggaacctataaaccgtatgggtccaacttt
YGR144W-50_LNA3	GgtAtgGaaGttGctGaamCtgGatGgaTtaAacmCgtAtgGgtmCcaActTt
YGR144W-50_LNA3_mt1	GgtAtgGaaGttGctGaamCtgGatmCgaTtaAacmCgtAtgGgtmCcaActTt
YGR144W-50_LNA3_mt2	GgtAtgGaaGttGctGaamCtgGatmCcaTtaAacmCgtAtgGgtmCcaActTt
YGR144W-50_LNA3_mt3	GgtAtgGaaGttGctGaamCtgGaamCcaTtaAacmCgtAtgGgtmCcaActTt
YGR144W-50_LNA3_mt4	GgtAtgGaaGttGctGaamCtgGaamCctTtaAacmCgtAtgGgtmCcaActTt
YGR144W-50_LNA3_mt5	GgtAtgGaaGttGctGaamCtgGaamCctAtaAacmCgtAtgGgtmCcaActTt

Example 2: Detection of Alternative Splice Isoforms Using Exon-Specific, Internal LNA Capture Probes in the *Caenorhabditis elegans* Gene *let-2*

*Capture probe design*

5 Finding the regions of interest

From the database “intronerator” [<http://www.cse.ucsc.edu/~kent/intronerator/>] as well as scientific literature, the *C. elegans* *Let-2* gene encoding type IV collagen was found according to the following criteria. The generation of mature mRNA desirably involves either complete exon or intron skipping. ESTs (expressed sequence tags) desirably indicate different 10 isoforms. If ESTs were only different from the gene annotation(s), this could simply mean that the prediction is wrong, and nothing more. Desirably, there are different EST splice indications in different developmental stages. Two gene prediction algorithms (e.g., GeneFinder and Genie) desirably agree upon the number of genes in a coding segment. Exons of interest (e.g., exons being skipped and their flanking exons) in the *C. elegans* gene T01D3.3 desirably exceed 15 70 base-pairs. Other genes of interest may be selected using one or more of the above criteria or using other criteria, such as the medical relevance of the gene.



### Determining melting temperatures and palindromic properties of the *C. elegans* let-2 gene/exons 8, 9, 10, and 11-specific capture probes

The script PICK70 (which was kindly provided by Jingchun Zhu from the Joe DeRisi Laboratory and which is publicly available) was used to run a sliding 50 base-pair window across the regions in which an oligonucleotide capture probe should be designed. The output data were saved for later.

### Determining the uniqueness of the regions

All regions were compared using a publicly available BLAST program to the complete set of annotated transcripts from the *C. elegans* genome downloaded from NCBI. For each region a list with the location of all BLAST hits was made.

### Choice of Desirable $T_m$ for Capture Probes

From the PICK70 output, the distribution of melting temperatures for all possible oligonucleotides was collected. As these centered around approximately 80°C, this temperature was chosen as the desirable temperature.

For each region, an oligonucleotide with a palindromic value below 100 (default value in PICK70, value based on Smith-Waterman algorithm) and with a melting temperature closest to 80°C was picked. The location of the oligonucleotide within the region was then compared to the list made using the above BLAST search. If the oligonucleotide did not coincide with a BLAST hit exceeding around 25 (consecutive) base-pairs, this oligonucleotide sequence was chosen as a 50-mer capture probe. Otherwise, a new oligo sequence was picked from the PICK70 output.

### Checking Probe Sequences

The selected 50-mer oligonucleotide sequences were "BLASTed" against the *C. elegans* transcripts again, as described above.

### Accounting for the Introns.

The oligonucleotide sequences were "BLASTed" against the complete *C. elegans* genome. The matches were run against a list made from the GenBank reports of the complete genome, indexing whether positions in the genome were genic or intergenic.

It was checked to determine whether new hits to genic regions appeared (compared to the initial BLAST search using the PICK70 output). If this was not case, the oligonucleotide sequences were selected for capture probe synthesis.

### Design of the LNA-modified Capture Probes

For the LNA-modified oligonucleotide capture probes, every fourth DNA nucleotide was substituted with an LNA nucleotide, as shown in Table 3. The oligonucleotides were synthesized with an anthraquinone (AQ) moiety at the 5'-end of each oligonucleotide (e.g., as

described in allowed U.S.S.N. 09/611,833), followed by a hexaethyleneglycol tetramer linker region and the LNA/DNA mixmer capture oligonucleotide sequence.

Table 3. *C. elegans* let-2 gene/exons 8, 9, 10, and 11-specific capture probes

5

Capture probe	Sequence (LNA=uppercase, DNA=lowercase letters)
CE42.08-0HEG4	<u>GgctGgatmCcccAggaAaccmCaggAatcGgaaGcatTggamCcaaAaggAg</u>
CE42.09-0HEG4	<u>mCaccGgatmCcggmCtcaAttgTcggAcctmCgcgGaaamCcctGgagAaaaGg</u>
CE42.10-0HEG4	<u>TccgmCcagGcccAatcGcctmCcacmCatgTccaAgggAaccAttaTcggTc</u>
CE42.11-0HEG4	<u>GagcmCaggAgagGgagGtcaAcgcGgttAcccAggaAatgGaggActcTc</u>

### Strains and Growth Conditions

*C. elegans* wild-type strain (Bristol-N2) was maintained on nematode growth medium (NG) plates seeded with *Escherichia coli* strain OP50 at 20°C, and the eggs and L1 larvae were prepared as described in Hope, I. A. (ed.) " *C. elegans* – A Practical Approach ", Oxford University Press 1999. The samples were immediately flash frozen in liquid N<sub>2</sub> and stored at – 80°C until RNA isolation.

### Isolation of Total RNA

A 100 µl aliquot of packed *C. elegans* worms from a L1 larvae population was homogenized using the FastPrep Bio101 from Kem-En-Tec for 1 minute at speed 6 followed by isolation of total RNA from the extracts using the FastPrep Bio101 kit (Kem-En-Tec) according to the manufacturer's instructions. A 50 µl aliquot of packed *C. elegans* eggs was homogenized in lysis buffer (RNeasy total RNA purification kit, QIAGEN) containing quartz sand for 3 minutes using a Pellet Pestle Motor followed by isolation of total RNA according to the manufacturer's manual.

The eluted total RNA from worms (L1 larvae) as well as eggs was ethanol precipitated for 24 hours at – 20°C by addition of 2.5 volumes of 96% EtOH and 0.1 volume of 3M Na-acetate, pH 5.2 (Ambion, USA), followed by centrifugation of the total RNA sample for 30 minutes at 13200 rpm. The total RNA pellet was air-dried and redissolved in 6 µl (worms) or 25 2.5 µl (eggs) of diethylpyrocarbonate (DEPC)-treated water (Ambion, USA) and stored at – 80°C.

### Reverse transcription (RT)-PCR

Total RNA (1.5 µg) from eggs or 1 µg total RNA from worms (L1 larvae) were mixed with 5 µg oligo(dT)12-18 primer (Amersham Pharmacia Biotech, USA) and 0.5 µg of random

hexamers, pd(N)<sub>6</sub> (Amersham Pharmacia Biotech, USA) and DEPC-treated water to a final volume of 7 µl. The mixture was heated at 70°C for 10 minutes, quenched on ice for 5 minutes, followed by addition of 20 units of Supersasin RNase inhibitor (Ambion, USA), 4 µl of 5 x Superscript buffer (Life Technologies, USA), 2 µl of 100 mM DTT, 1 µl of dNTP solution 5 (20 mM each dATP, dGTP, dTTP and dCTP, Amersham Pharmacia Biotech, USA), and 3 µl of DEPC-treated water.

The primers were pre-annealed at 37°C for 5 minutes, followed by addition of 400 units of Superscript II reverse transcriptase (Invitrogen, USA). First strand cDNA synthesis was carried out at 37°C for 30 minutes, followed by 2 hours at 42°C, and the reaction was stopped by incubation at 70°C for 5 minutes, followed by incubation on ice for 5 minutes.

Unincorporated dNTPs were removed by gel filtration using MicroSpin S-400 HR columns as described below. The column was pre-spun for 1 minute at 735 x g in a 1.5 ml tube, and the column was placed in a new 1.5 ml tube. The cDNA sample was slowly applied to the top center of the resin and spun at 735 x g for 2 minutes. The eluate was collected. The 15 volume of the eluate was adjusted to 50 µl with TE-buffer pH 7.0 before being used as the template for linear PCR. Four µl template (RT from eggs or worms) was combined with 1 µl dNTP solution (10mM each dATP, dGTP, dTTP and dCTP, Amersham Pharmacia Biotech, USA), 1 µl of each primer (20µM CE42.07 sense: gatcgaattcctccaggagagaagggagatg, and CE42.12 antisense: 5' gatcaagcttatctcttctgggtatccagctt), 5 µl 10 x AmpliTaq Gold Polymerase 20 buffer, 5µl 25 mM MgCl<sub>2</sub>, 0.5 µl AmpliTaq Gold DNA polymerase (5U/µl, Applied Biosystems), 2 µl Cy3-dCTP (Amersham Pharmacia Biotech, USA) (eggs) or 2 µl Cy5-dCTP (Amersham Pharmacia Biotech, USA) (worms), and 31.5 µl DEPC-treated water to a final volume of 50 µl. The PCR reactions were carried out using the following program: 95°C for 5 minutes followed by 30 cycles of PCR using the following cycling program (denaturation at 95°C for 45 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 1 minute) followed by a final extension step at 72°C for 10 minutes and incubation on ice for 5 minutes.

Purification of the PCR amplicons from eggs as well as worms was performed using a Qiaquick PCR purification kit (QIAGEN) according to the manufacturer's instructions.

#### *Fluorochrome-labeling of the let-2 cDNA Fragments using Primer Extension*

30 Four (4) µl template (RT from eggs or worms) was combined with 1 µl dNTP solution (10mM each dATP, dGTP, dTTP and dCTP, Amersham Pharmacia Biotech, USA), 1 µl of each primer (20µM CE42.12 antisense 5' gatcaagcttatctcttctgggtatccagctt), 5 µl 10 x AmpliTaq Gold Polymerase buffer, 5µl 25 mM MgCl<sub>2</sub>, 0.5 µl AmpliTaq Gold DNA polymerase (5U/µl, Applied Biosystems), 2 µl Cy3-dCTP (Amersham Pharmacia Biotech, USA) (eggs) or 2 µl Cy5-

dCTP (Amersham Pharmacia Biotech, USA) (worms), and 31.5 µl DEPC-treated water to a final volume of 50 µl. The PCR reactions were carried out using the following program: 95°C for 5 minutes followed by 30 cycles of PCR using the following cycling program (denaturation at 95°C for 45 seconds annealing at 60°C for 30 seconds extension at 72°C for 1 minute) followed by a final extension step at 72°C for 10 minutes and incubation on ice for 5 minutes.

Purification of the PCR amplicons from eggs as well as worms were performed using a Qiaquick PCR purification kit (QIAGEN) according to the manufacturer's instructions. Unincorporated dNTP nucleotides were removed by gel filtration using MicroSpin S-400 HR columns as described above before the eluted, fluorochrome-labelled DNA fragments were stored at -20°C in the dark until microarray hybridization.

#### *Printing and Coupling of the C. elegans Let-2 Exon 8-11 Microarrays*

The *C. elegans* gene Let-2/exon 8-11 capture probes were synthesized with a 5' anthraquinone (AQ)-modification, followed by a hexaethyleneglycol-4 (HEG4) linker (Table 3). The capture probes were first diluted to a 10 µM final concentration in 100 mM Na<sub>2</sub>HPO<sub>4</sub> pH 7.0 and spotted on Euray COP microarray slides using the Biochip Arrayer One (Packard Biochip Technologies) with a spot volume of 300 pl and 300 µm between the spots.

The capture probes were immobilized onto the microarray slide by UV irradiation in a Stratagene for 90 seconds at full power (Stratagene, USA). Non-immobilized capture probe oligonucleotides were removed from the slides by washing the slides for ½ hour in 30% acetone before rising in milli-Q H<sub>2</sub>O. After washing, the slides were centrifuged at 800 rpm for 2 minutes and stored in a slide box until microarray hybridization.

#### *Comparative Hybridization of the C. elegans microarrays and Post-hybridization Washes*

The slides were hybridized with 2.5 µl of the Cy3-labelled and 2.5 µl of the Cy5-labelled target preparation from eggs and worms, respectively, as described above (see "Reverse transcription (RT)-PCR" section) in 25 µl of hybridization solution, containing 25 mM HEPES, pH 7.0, 3 x SSC, 0.3% SDS, and 25 µg of yeast tRNA. The target probe was filtered in a Millipore 0.22 micron spin column (Ultrafree-MC, Millipore, USA), denatured by incubation at 100°C for 5 minutes, cooled at room temperature for 5 minutes, and then carefully applied onto the prepared microarray. One-third of a cover slip was laid over the microarray, and the hybridization was performed for 16-18 hours at 65°C in a hybridization chamber (DieTech, model Joe deRisi, USA).

Following hybridization, the slides were washed sequentially by plunging gently in 2 x SSC/0.1% SDS at room temperature until the cover slip falls off into the washing solution, then

in 1x SSC pH 7.0 (150 mM NaCl, 15 mM Sodium Citrate) at room temperature for 1 minute, then in 0.2 x SSC, pH 7.0 (30 mM NaCl, 3 mM Sodium Citrate) at room temperature for 1 minute, and finally in 0.05 x SSC (7.5 mM NaCl, 0.75 mM Sodium Citrate) for 5 seconds, followed by drying of the slides by spinning at 500 rpm for 5 minutes. The slides were stored in a slide box in the dark until scanning.

#### *Microarray data analysis*

The *C. elegans let-2* gene microarray was scanned in an ArrayWoRx Scanner (Applied Precision, USA) using an exposure time of 5 seconds, resolution of 5.0, and high (high level) sensitivity. The hybridization data were analyzed using the ArrayVision image analysis software package 5.1 (IMAGING Research Inc., USA). The detection principle for alternative exon skipping the *C. elegans let-2* gene is shown in Fig. 14. As demonstrated in Fig. 15, analysis of the comparative hybridization data from the *C. elegans Let-2* exon 8-11 array demonstrates detection of alternative exon skipping of the *let-2* exon 9 (eggs) and exon 10 (L1 larvae) using LNA-modified 50-mer capture probes. Capture probes of other lengths and/or with other LNA substitution patterns can be used similarly.

#### Example 3: Improved Sensitivity in the Specific Detection of the *C. elegans* Gene T01D3.3 Exon 4 Using LNA-Modified Oligonucleotide Capture Probes

**Capture Probe Design:** The design method of exon-specific capture probes for the *C. elegans* gene T01D3.3 exon 4 has been described in example 2.

**Design of the LNA-modified Capture Probes:** For the LNA-spiked oligonucleotide capture probes, every fourth DNA nucleotide was substituted with an LNA nucleotide, as shown in Table 4.

Table 4. *C. elegans* gene T01D3.3/exon 4-specific capture probes.

Capture probes	Sequence (LNA=uppercase, DNA=lowercase letters)
CEgene26.04-70	ggctggaacagaagttgttggtgcgtgacaaggtatggaagaagattatccggaaaagaaagcaa agac
CEgene26.04-50	ggctggaacagaagttgttggtgcgtgacaaggtatggaagaagattat
CEgene26.04-40	ggctggaacagaagttgttggtgcgtgacaaggtatgga
CEgene26.04-30	gaacagaagttgttggtgcgtgacaaggt
CEgene26.04-50HEG2	GgctGgaamCagaAgtTgttGgtgmCgtgAcaaGgtaTggaAgaaGattAt
CEgene26.04-50HEG4	GgctGgaamCagaAgtTgttGgtgmCgtgAcaaGgtaTggaAgaaGattAt
CEgene26.04-40HEG2	GgctGgaamCagaAgtTgttGgtgmCgtgAcaaGgtaTgga
CEgene26.04-40HEG4	GgctGgaamCagaAgtTgttGgtgmCgtgAcaaGgtaTgga
CEgene26.04-30HEG2	GaacAgaaGtttGttGgtcGtgamCaagGt
CEgene26.04-30HEG4	GaacAgaaGtttGttGgtcGtgamCaagGt

### *Cultivation of Caenorhabditis elegans Worms*

Mixed stage *C. elegans* cultures were grown according to standard methods. Samples were harvested by centrifugation at 3000×g, suspended in RNA Later (Ambion, USA), and immediately frozen in liquid nitrogen.

### *mRNA Isolation from C. elegans Mixed Stages Worms*

Poly(A)<sup>+</sup>RNA was isolated from the worm samples using the Pick-Pen (Bio-Nobile, Finland) Starter kit combined with the KingFisher mRNA purification kit (ThermoLabsystems, Finland) according to the manufacturer's instructions. The yield was 1-2 µg poly(A)<sup>+</sup>RNA from approximately 50 mg of *C. elegans* worms.

### *Synthesis of fluorochrome labelled first strand cDNA from C. elegans mRNA*

One µg of *C. elegans* poly(A)<sup>+</sup>RNA was combined with 2 µg oligo dT primer (T20VN) in an RNase free, pre-siliconized 1.5 mL tube, and the final volume was adjusted with DEPC-water to 8 µL. The reaction mixture was heated at +70°C for 10 minutes, quenched on ice 15 minutes, spun for 20 seconds, followed by addition of 1 µL SUPERase-In™ (20U/µL, RNase inhibitor, Ambion, USA), 4 µL 5×RTase buffer (Invitrogen, USA), 2 µL 0.1 M DTT (Invitrogen, USA), 1 µL dNTP (20mM dATP, dGTP, dTTP; 4 mM dCTP in DEPC-water, Amersham Pharmacia Biotech, USA), and 3 µL Cy3™-dCTP (Amersham Pharmacia Biotech, USA). First strand cDNA synthesis was carried out by adding 1 µL of Superscript™ II (Invitrogen, 200 U/mL), mixing, and incubating the reaction mixture for one hour at 42°C. An additional 1 µL of Superscript™ II was added and the cDNA synthesis reaction mixture was incubated for an additional one hour at 42°C; the reaction was stopped by heating at 70°C for 5 minutes, and quenching on ice for 2 minutes. The RNA was hydrolyzed by adding 3 µL of 0.5 M NaOH and incubating at 70°C for 15 minutes. The samples were neutralized by adding 3 µL of 0.5 M HCl and purified by adding 450 µL 1×TE buffer, pH 7.5 to the neutralized sample and transferring the samples onto a Microcon-30 concentrator. The samples were centrifuged at 14000×g in a microcentrifuge for ~8 minutes, the flow-through was discarded, and the washing step was repeated twice by refilling the filter with 450 µl 1×TE buffer and by spinning for ~12 minutes. Centrifugation was continued until the volume was reduced to 5 µL, and finally the 30 labelled cDNA probe was eluted by inverting the Microcon-30 tube and spinning at 1000×g for 3 minutes.

### *Printing and Coupling of the C. elegans Microarrays*

The *C. elegans* gene T01D3.3/exon 4 capture probes were synthesized with a 5' anthraquinone (AQ)-modification, followed by either a hexaethyleneglycol-2 or a

hexaethyleneglycol-4 (HEG2/HEG4) linker (Table 4). The capture probes were first diluted to a 10  $\mu$ M final concentration in 100 mM Na-phosphate buffer pH 7.0, followed by a two-fold dilution series (10  $\mu$ M, 5  $\mu$ M, 2.5  $\mu$ M, 1.25  $\mu$ M, 0.625  $\mu$ M, 0.31  $\mu$ M, and 0.155  $\mu$ M) and spotted on Exiqon's polycarbonate microarray slides using the Biochip Arrayer One (Packard 5 Biochip Technologies, USA) with a spot volume of 3x 300 pl and 400  $\mu$ m between the spots. The capture probes were immobilized onto the microarray slide by UV irradiation in a Stratalinker for 90 seconds at full power (Stratagene, USA). Non-immobilized capture probe oligonucleotides were removed from the slides by washing the slides for 24 hours in milli-Q H<sub>2</sub>O. After washing, the slides were dried in an oven at 37°C for 30 minutes and stored in a 10 slide box until microarray hybridization.

#### *Hybridization with Cy3-labelled cDNA*

The arrays were hybridized overnight using the following protocol. The Cy3<sup>TM</sup>-labelled cDNA sample was combined with 3  $\mu$ L 20 $\times$ SSC (3 $\times$ SSC final), 0.5  $\mu$ L 1 M HEPES, pH 7.0 (25 mM final), 25  $\mu$ g yeast tRNA (1.25  $\mu$ g/ $\mu$ L final), 10  $\mu$ g PolyA blocker (0.5  $\mu$ g/ $\mu$ L final), 0.6 15  $\mu$ L 10% SDS (0.3% final), and DEPC-treated water to 20  $\mu$ L final volume. The labelled cDNA target sample was filtered in a Millipore 0.22 micron spin column according to the manufacturer's instructions (Millipore, USA), and the probe was denatured by incubating the reaction at 100°C for 2 minutes. The sample was cooled at 20-25°C for 5 minutes by spinning at maxium speed in a microcentrifuge, and then carefully applied on top of the microarray. A 20 cover slip was laid over the microarray and the hybridization was performed for 16 hours at 63°C in a hybridization chamber (Corning, USA) submerged in a water bath, with an aliquot of 30  $\mu$ L of 3 $\times$ SSC added to both ends of the hybridization chamber to prevent evaporation. After hybridization, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The coverslip was washed off in 2 x SSC, pH 7.0 containing 25 0.1% SDS at room temperature for one minute, followed by washing of the microarrays subsequently in 1.0 x SSC, pH 7.0 at room temperature for one minute, and then in 0.2 x SSC, pH 7.0 at room temperature for one minute. Finally, the slides were washed for 5 seconds in 0.05 x SSC, pH 7.0. The slides were then dried by centrifugation in a swinging bucket rotor at approximately 600 rpm for 5 minutes.

#### 30 *Microarray data analysis*

The *C. elegans* gene T01D3.3 exon 4 array was scanned in an ArrayWoRx Scanner (Applied Precision, USA) using an exposure time of 5 seconds, resolution of 5.0, and high (high level) sensitivity. The hybridization data were analyzed using the ArrayVision image analysis software package 5.1 (IMAGING Research Inc., USA). As shown in Figs. 16A and

16B, analysis of the hybridization data from the *C. elegans* gene 26/T01D3.3 exon 4 array demonstrates that the use of LNA-modified capture probes for the *C. elegans* T01D3.3 exon results in 5-fold increased sensitivity in exon 4 capture compared to the corresponding DNA oligonucleotide capture probe controls printed on the same microarray. Capture probes of other 5 lengths and/or with other LNA substitution patterns can be used similarly.

**Example 4: Assessment of Capture Probe Specificity for the *C. elegans* Gene T01D3.3 Exons 4 and 5 Using Synthetic Antisense Target Oligos**

*Capture probe design:* Exon-specific capture probes for the *C. elegans* gene T01D3.3 exons 4 and 5 were designed as described in Example 2.

*Design of the LNA-modified capture probes:* For the LNA-spiked oligonucleotide capture probes, every fourth DNA nucleotide was substituted with an LNA nucleotide, as shown in Table 5: *C. elegans* gene T01D3.3/exons 4 and 5-specific capture probes and synthetic target oligonucleotides.

15 Table 5.

<b>Capture probes</b>	<b>Sequence (LNA=uppercase, DNA=lowercase letters)</b>
CEgene26.04-70	ggctggaacagaagttgttggtgctgacaagggtatggaagaagattatccggaagaa agcaaagac
CEgene26.05-70	tatgtggcgcaatgagcaatattcagcatgtttctcctctgtcaaccatcatgtcaagatcctt caac
CEgene26.04-50	ggctggaacagaagttgttggtgctgacaagggtatggaagaagattat
CEgene26.05-50	tatgtggcgcaatgagcaatattcagcatgtttctcctctgtcaacca
CEgene26.04-40	ggctggaacagaagttgttggtgctgacaagggtatgga
CEgene26.05-40	tatgtggcgcaatgagcaatattcagcatgtttctcctc
CEgene26.04-30	gaacagaagttgttggtgctgacaagggt
CEgene26.05-30	tatgtggcgcaatgagcaatattcagcat
CEgene26.04-50HEG2	GgctGgaamCagaAgttTgtTGgtgmCgtgAcaaGgtaTggaAgaaGattAt
CEgene26.04-50HEG4	GgctGgaamCagaAgttTgtTGgtgmCgtgAcaaGgtaTggaAgaaGattAt
CEgene26.05-50HEG2	TatGTggcGcgaAtgaGcaaTattmCagcAtgtTtctmCctcTtgTmCaacmCa
CEgene26.05-50HEG4	TatGTggcGcgaAtgaGcaaTattmCagcAtgtTtctmCctcTtgTmCaacmCa
CEgene26.04-40HEG2	GgctGgaamCagaAgttTgtTGgtgmCgtgAcaaGgtaTgga
CEgene26.04-40HEG4	GgctGgaamCagaAgttTgtTGgtgmCgtgAcaaGgtaTgga
CEgene26.05-40HEG2	TatGTggcGcgaAtgaGcaaTattmCagcAtgtTtctmCctc
CEgene26.05-40HEG4	TatGTggcGcgaAtgaGcaaTattmCagcAtgtTtctmCctc
CEgene26.04-30HEG2	GaacAgaaGtttGttgGtgcGtgamCaagGt
CEgene26.04-30HEG4	GaacAgaaGtttGttgGtgcGtgamCaagGt
CEgene26.05-30HEG2	TatGTggcGcgaAtgaGcaaTattmCagcAt
CEgene26.05-30HEG4	TatGTggcGcgaAtgaGcaaTattmCagcAt



<b>Target oligos</b>	<b>Sequence (LNA=uppercase, DNA=lowercase letters)</b>
CEgene26.04-biotarget	accctgtcacgcaccaaaactctgttc
CEgene26.05-biotarget	atgctgaatattgctcattcgccacata

*Printing and coupling of the C. elegans gene T01D3.3/exon 4-5 microarrays*

The *C. elegans* gene T01D3.3/exon 4-5 capture probes were synthesized with a 5' anthraquinone (AQ)-modification, followed by either a hexaethyleneglycol-2 or a hexaethyleneglycol-4 (HEG2/HEG4) linker (Table 5). The capture probes were first diluted to a 10  $\mu$ M final concentration in 100 mM Na-phosphate buffer pH 7.0, followed by a two-fold dilution series (10  $\mu$ M, 5  $\mu$ M, 2.5  $\mu$ M, 1.25  $\mu$ M, 0.625  $\mu$ M, 0.31  $\mu$ M, and 0.155  $\mu$ M) and spotted on Euray polycarbonate microarray slides using the Biochip Arrayer One (Packard 10 Biochip Technologies) with a spot volume of 3x 300 pl and 400  $\mu$ m between the spots. The capture probes were immobilized onto the microarray slide by UV irradiation in a Stratalinker for 90 seconds at full power (Stratagene, USA). Non-immobilized capture probe oligonucleotides were removed from the slides by washing the slides for 24 hours in milli-Q H<sub>2</sub>O. After washing, the slides were dried in an oven at 37°C for 30 minutes, and stored in a 15 slide box until microarray hybridization.

*Hybridization of the C. elegans microarrays and post-hybridization washes*

The slides were hybridized with a high (saturated) concentration of 1 $\mu$ M of each gene T01D3.3, exon 4 or 5 target oligo (Table 5) in 50  $\mu$ l of hybridization solution, containing 25 mM HEPES, pH 7.0, 3 x SSC, 0.22 % SDS, and 0.8  $\mu$ g/ $\mu$ l of poly(A) blocker. The target 20 probes were filtered in a Millipore 0.45 micron spin column (Ultrafree-MC, Millipore, USA), denatured by incubation at 100 °C for 2 minutes, cooled at room temperature for 5 minutes, and then carefully applied onto the prepared microarray. One-half of a cover slip was laid over the microarray, and the hybridization was performed for 16-18 hours at 63°C in a hybridization chamber (Corning, USA).

25 Following hybridization, the slides were washed sequentially by plunging gently in 1 x SSCT (150 mM NaCl, 15 mM Sodium Citrate + Tween 20) at room temperature for one minute, then in 0.2 x SSCT (30 mM NaCl, 3 mM Sodium Citrate + Tween 20) at room temperature for one minute, and finally in Milli Q water, followed by drying of the slides in an oven at 37°C for 30 minutes. The slides were Cy5 labelled using a Cy5-streptavidin target. 30 Thirty  $\mu$ l of a Cy5-streptavidin (2 $\mu$ g/ml in 1 x SSCT) were carefully applied onto the hybridized microarray and incubated one hour at room temperature before an additional washing step were performed in 1 x SSCT (150 mM NaCl, 15 mM Sodium Citrate + Tween 20)

at room temperature for one minute, then in 0.2 x SSCT (30 mM NaCl, 3 mM Sodium Citrate + Tween 20) at room temperature for one minute, and finally in Milli Q water. Following washing, the slides were drying in an oven at 37°C for 30 minutes and stored in a slide box in the dark until scanning.

### 5 Microarray data analysis

The *C. elegans* gene T01D3.3 exon 4-5 microarray was scanned in an ArrayWoRx Scanner (Applied Precision, USA) using an exposure time of 5 seconds, resolution of 5.0, and high (high level) sensitivity. The hybridization data were analyzed using the ArrayVision image analysis software package 5.1 (IMAGING Research Inc., USA). As shown in Figs. 18A and 18B, analysis of the hybridization data from the *C. elegans* gene 26/T01D3.3 array demonstrates that both the DNA as well as the LNA capture probes for the *C. elegans* T01D3.3 exons 4 (Fig. 18A) and exon 5 (Fig. 18B), respectively are highly specific with a very low level of cross-hybridization between their respective target oligonucleotides. The exon-specific design of the oligonucleotide capture probes is thus validated. Capture probes of other lengths 15 and/or with other LNA substitution patterns can be used similarly.

### Example 5: Detection of Alternatively Spliced Isoforms using Internal Exon-specific, and Exon-Exon Junction-Specific (merged) LNA-modified Capture Probes

#### *Oligonucleotide design for microarrays.*

20 Methods for designing exon-specific internal oligonucleotide capture probes has been described in Example 2.

#### *Design of the LNA-modified capture probes*

For the LNA-modified oligonucleotide capture probes, every third DNA nucleotide was substituted with an LNA nucleotide. The probes designed to capture the junction of the 25 recombinant splice variants were designed with LNA modifications in a block of five consecutive LNAs nucleotides, two on the 5' side of the splice junction and three on the 3' side of the splice junction. All capture probes are shown in Table 6.

Table 6. Internal, exon-specific and merged, exon-exon junction specific oligonucleotide 30 capture probes.

Capture probes	Sequence (LNA=uppercase, DNA lowercase letters)
gene78.01a	cctgaaagtagatttgatttccgaaacgccttctcccgttcttaagtc
gene78.01b	catataccacaataatgtccctcaaaaatcacagaaaactcacaacactg
gene78.03a	gatttgcagcgggtgtaaaaagtatgaaacgtggaattaaaaggctc
gene78.03b	ccaatgaaaactaatcaaggtaaacgtggatcccatggcaattcccggg

gene78.m01INS3	caacactgccagaggttcaatcgatccgatcctaagaaggcgccc
gene78.mINS303	gtccagatcgtccatcatagatcgataaatgtgaaggaaatgcctg
gene78.m01INS4	caacactgccagaggttcaatcgatgtgtgataggatcagtggtcagg
gene78.mINS403	gaaggcgaaggagactgctaatacgataaatgtgaaggaaatgcctg
gene78.01a_50_LNA3	mCctGaaAgtAgaTttGttAttTccGaaAcgmCctTctmCccGttmCttAagTc
gene78.01b_50_LNA3	mCatAtamCcamCaaAtaGtcmCctmCaaAaaTcamCaaGaaAacTcamCaamCacTg
gene78.03a_50_LNA3	GatTtgmCagmCggTggTaaAaaGtaTgaAaamCgtGgtAatTaaAagGtcTc
gene78.03b_50_LNA3	mCcaAtgAaaActAatmCaaAggTaaAcgTggAtcmCcaTggmCaaTtcmCcgGg
gene78.m01INS3_50_block	caacactgccagaggttcaatcGATmCmCgatgatcctaagaaggcgccc
gene78.mINS303_50_block	gtccagatcgtccatcatAGTATcgataaatgtgaaggaaatgcctg
gene78.m01INS4_50_block	caacactgccagaggttcaatcGATGTgtgataggatcagtggtcagg
gene78.mINS403_50_block	gaaggcgaaggagactgctAATATcgataaatgtgaaggaaatgcctg

### *Printing and Coupling of the Splice Isoform-Specific Microarrays*

The splice variant capture probes were synthesized with a 5' anthraquinone (AQ)-5 modification, followed by a hexaethyleneglycol-2 (HEG2) linker. The capture probes were first diluted to a 20  $\mu$ M final concentration in 100 mM Na-phosphate buffer pH 7.0, and spotted on the Immobilizer polymer microarray slides (Exiqon, Denmark) using the Biochip Arrayer One (Packard Biochip Technologies, USA) with a spot volume of 2x 300 pl and 300  $\mu$ m between the spots. The capture probes were immobilized onto the microarray slide by UV irradiation in 10 a Stratalinker with 2300  $\mu$ joules (Stratagene, USA). Non-immobilized capture probe oligonucleotides were removed from the slides by washing the slides two times 15 minutes in 1xSSC. After washing, the slides were dried by centrifugation at 1000x g for 2 minutes, and stored in a slide box until microarray hybridization.

### *Construction of Splice Variant Vectors*

15 The recombinant splice variant constructs were cloned into the Triamp18 vector (Ambion, USA). The constructs were sequenced to confirm their construction. The plasmid clones were transformed into *E. coli* XL10-Gold (Stratagene, USA).

### *Triamp18/SWI5 vector construct*

20 Genomic DNA was prepared from a wild-type standard laboratory strain of *Saccharomyces cerevisiae* using the Nucleon MiY DNA extraction kit (Amersham Biosciences, USA) according to the supplier's instructions. Amplification of the partial yeast gene was performed using standard PCR with yeast genomic DNA as the template. In the first step of amplification, a forward primer containing a restriction enzyme site and a reverse primer containing a universal linker sequence were used. In this step, 20 base-pairs were added to the

3'-end of the amplicon, next to the stop codon. In the second step of amplification, the reverse primer was exchanged with a nested primer containing a poly-T<sub>20</sub> tail and a restriction enzyme site. The *SWI5* amplicon contains 730 bp of the *SWI5* ORF plus a 20 bp universal linker sequence and a poly-A<sub>20</sub> tail. The PCR primers used were YDR146C-For-EcoRI 5 (acgtgaattcaatacagacaatgaaggagatga),

YDR146C-Rev-Uni (gatccccgggaattgccatgttacctttgattagttttcattggc), and Uni-polyT-BamHI (acgtggatcctttttttttttttttttgatccccgggaattgccatg).

The PCR amplicon was cleaved with the restriction enzymes *EcoRI* and *BamHI*. The DNA fragment was ligated into the pTRIamp18 vector (Ambion, USA) using the Quick 10 Ligation Kit (New England Biolabs, USA) according to the supplier's instructions and transformed into *E. coli* DH-5 $\alpha$  by standard methods.

#### *Construction of the Recombinant Splice Variant #1 (Triamp18/swi5-rubisco)*

The *Arabidopsis thaliana* Rubisco small subunit ssu2b gene fragment (gi17064721) was amplified from genomic DNA using primers named DJ305 (5'-15 ACTATGATGGACGATACTGGAC-3') and DJ306 (5'-ATTGGATCGATCCGATGATCCTAATGAAGGC-3'), containing *ClaI* restriction site linkers. The purified PCR fragment was digested with *ClaI* and then cloned into the swi5 (gi:7839148) vector at the unique *ClaI* site (atcgat) giving each insert a flanking sequence from the original yeast *SWI5* insert (named exon01 and exon 03, Fig. 19). The product was inserted 20 in the reverse orientation, so that the insert sequence is as follows:

atcgatCCGATGATCCTAATGAAGGCGCCCGGGTACTCCTTCTTGCAATTCTTCAACTTC  
CTTCAACACTTGAGCGGAGTCGGTGCATCCGAACAATGGAAGCTTCCACATTGTCC  
AGTATCGTCCATCATAGTatcgat

Nucleotide sequence analysis revealed a difference between the sequence of *A. thaliana* 25 rubisco expected from the GenBank database and that obtained from all sequenced constructs and PCR products. Position 30 in the Rubisco insert is "C" rather than the expected "A." This SNP was probably created by PCR. None of the oligonucleotide capture probes used in the example cover this region. The Rubisco sequence in Genbank is TCCTAATGAAGGCGCCA, and the sequence obtained from the plasmid construct is TCCTAATGAAGGCGCCC.

#### 30 *Construction of the Recombinant Splice Variant # 2 (Triamp18/swi5-lea)*

The *Arabidopsis thaliana* Lea gene (gi1526423) was amplified from genomic DNA with primers named DJ307 (5'-GGAATTATCGATGTGTGATAGGATCAGTGTTTCAG-3'), and DJ308 (5'-AATTGGATCGATATTAGCAGTCTCCTTCGCC-3'), including the *ClaI* linker sites as above. The PCR fragment was digested with *ClaI* cloned into the yeast *SWI5* 35 IVT construct as above at the unique *ClaI* site.

The fragment was inserted in the forward orientation, resulting in the following insert sequence:

atcgatGTGTGATAGGTTTCAGTGTTTCAGGGCTGTCCAAGGAACGTATGAGCATGCGAG  
AGACGCTGTAGTTGGAAAAACCCACGAAGCGGCTGAGTCTACCAAAGAAGGAGCT  
5 CAGATAGCTTCAGAGAAAGCGGTTGGAGCAAAGGACGCAACCGTCGAGAAAGCTA  
AGGAAACCGCTGATTATACTGCGGAGAAGGTGGGTGAGTATAAAGACTATACGGT  
TGATAAAGCTAAAGAGGCTAAGGACACAACCTGCAGAGAAGGCGAAGGAGACTGCT  
AATatcgat.

#### *In vitro RNA Preparation from Splice Variant Vectors*

- 10 *In vitro* RNA from the splice variants were made using the MEGAscript™ high yield transcription kit according to the manufacturer's instructions (Ambion, USA). The yield of IVT RNA was quantified at a Nanodrop spectrophotometer (Nanodrop Technologies, USA, Fig. 19).

#### *Isolation of total RNA from C.elegans*

- 15 *C. elegans* wild-type strain (Bristol-N2) was maintained on nematode growth medium (NG) plates seeded with *Escherichia coli* strain OP50 at 20°C, and the mixed stages of the nematode were prepared as described by Hope (ed.) ("*C. elegans* – A Practical Approach", Oxford University Press 1999). The samples were immediately flash frozen in liquid N<sub>2</sub> and stored at -80°C until RNA isolation.

- 20 A 100 µl aliquot of packed *C. elegans* worms from a mixed stage population was homogenized using the FastPrep Bio101 from Kem-En-Tec for one minute, speed 6 followed by isolation of total RNA from the extracts using the FastPrep Bio101 kit (Kem-En-Tec) according to the manufacturer's instructions.

The eluted total RNA was ethanol precipitated for 24 hours at -20°C by addition of 2.5 volumes of 96% EtOH and 0.1 volume of 3M Na-acetate, pH 5.2 (Ambion, USA), followed by centrifugation of the total RNA sample for 30 minutes at 13200 rpm. The total RNA pellet was air-dried and redissolved in 10 µl of diethylpyrocarbonate (DEPC)-treated water (Ambion, USA) and stored at -80°C.

#### *Fluorochrome-labelling of the Target*

- 30 Ten (10) µg total RNA from *C. elegans* and 1 ng of *in vitro* RNA from Splice variant #1 were combined with 5 µg anchored oligo(dT<sub>20</sub>) primer and DEPC-treated water to a final volume of 8 µl. The mixture was heated at 70°C for 10 minutes, quenched on ice for 5 minutes, followed by addition of 20 units of Supersasin RNase inhibitor (Ambion, USA), 1 µl dNTP solution (10mM each dATP, dGTP, dTTP and 0.4 mM dCTP, and 3 µl Cy5-dCTP,

Amersham Biosciences, USA), 4 µl 5 x RTase buffer (Invitrogen), 2µl 0.1 mM DTT (Invitrogen), 400 units of Superscript II reverse transcriptase (Invitrogen, USA), and DEPC-treated water to 20 µl final volume.

A parallel set-up was made with 10 µg total RNA from *C.elegans* and 1 ng of *in vitro* 5 RNA from Splice variant #2, labelling with Cy3-dCTP. Both cDNA syntheses were carried out at 42°C for 2 hours, and the reactions were stopped by incubation at 70°C for 5 minutes, followed by incubation on ice for 5 minutes.

Unincorporated dNTPs were removed by gel filtration using MicroSpin S-400 HR columns as described below. The column was pre-spun for one minute at 1500 x g in a 1.5 ml tube, and the column was placed in a new 1.5 ml tube. The cDNA sample was slowly applied to the top center of the resin and spun 1500-x g for 2 minutes. The eluate was collected. RNA was degraded by adding 3 µl of 0.5 M NaOH. The solution was mixed well and incubated at 70 °C for 15 minutes. The solution was neutralized by adding 3 µl of 0.5 M HCl and mixed well. Then, 450 µl 1xTE, pH 7.5 was added to the neutralized sample, and the sample was transferred 15 onto a Microcon-30 concentrator (prior to use, 500 µl 1xTE was spun through the column to remove residual glycerol). The samples were spun at 14000 x g in a micro centrifuge for 12 minutes, and the volume was checked. Spinning was continued until the volume was reduced to 5 µl. The labelled cDNA probe was eluted by inverting the Microcon-30 tube and spinning at 1000 x g for 3 minutes. The Microcon filter was checked for proper elution.

## 20 Comparative Hybridization of the Splice Variant Microarrays and Post-hybridization washes

The Cy3 and Cy5-labelled cDNA samples, respectively, were combined in one tube. The following was added: 3.75 µl 20x SSC (3x SSC final, pass through 0.22 µfilter prior to use to remove particulates), yeast tRNA (1 µg/µl final), 0.625 µl 1 M HEPES, pH 7.0 (25 mM final, pass through 0.22 µfilter prior to use to remove particulates), 0.75 µl 10 % SDS (0.3 % final), 25 and DEPC-water to 25 µl final volume. The labelled cDNA target sample was filtered in Millipore 0.22 µ filter spin column (Ultrafree-MC, Millipore, USA) according to the manufacturer's instructions, followed by incubation of the reaction mixture at 100 °C for 2-5 minutes. The cDNA probes were cooled at room temperature for 2-5 minutes by spinning at maximum speed in a microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was 30 carefully placed on top of the microarray spotted on Immobilizer™ MicroArray Slide, and the hybridization mixture was applied to the array from the side. An aliquot of 30 µL of 3xSSC was added to both ends of the hybridization chamber, and the Immobilizer™ MicroArray Slide was placed in the hybridization chamber (DieTech, USA). The chamber was sealed watertight and incubated at 65°C for 16-18 hours submerged in a water bath. After hybridization, the slide

was removed carefully from the hybridization chamber and washed using the following protocol.

The slides were washed sequentially by plunging gently in 2 x SSC/0.1% SDS at room temperature until the cover slip falls off into the washing solution, then in 1x SSC pH 7.0 (150 mM NaCl, 15 mM Sodium Citrate) at room temperature for one minute, then in 0.2 x SSC, pH 7.0 (30 mM NaCl, 3 mM Sodium Citrate) at room temperature for one minute, and finally in 0.05 x SSC (7.5 mM NaCl, 0.75 mM Sodium Citrate) for 5 seconds, followed by drying of the slides by spinning at 1000 x g for 2 minutes. The slides were stored in a slide box in the dark until scanning.

#### 10 Microarray data analysis

The splice variant microarray was scanned in a ScanArray 4000XL confocal laser scanner (Packard Instruments, USA). The hybridization data were analyzed using the GenePix Pro 4.01 microarray analysis software (Axon, USA).

In the data analysis, the experimental variation in the labelling efficiency between the two fluorescent dyes was normalized (scaled) as follows. The average signal intensities from the “exon1” and “exon3” internal capture probes (Table 6), were used to calculate normalization factor of 2.75. This factor was multiplied to the signal intensity values from the Cy-3 target.

Analysis of the data from the specific detection of the two recombinant splice variants in a complex RNA pool demonstrates that the merged capture probes containing a LNA block have significantly higher signals and a very low level of cross-hybridization, compared to the DNA capture probes (Figs. 20A and 20B). In addition, the specific detection of the two artificial splice variants #1 and #2 is validated with the results from LNA-modified oligonucleotide capture probes. Capture probes of other lengths and/or with other LNA substitution patterns can be used similarly. In contrast, the corresponding DNA oligonucleotide capture probes fail to detect splice variant #1 (Fig. 20B).

#### Example 6: The Use of LNA-modified Oligonucleotides in Microarrays Provides Significantly Improved Sensitivity in Expression Profiling

This example demonstrates the advantages of using LNA oligonucleotide microarrays in gene expression profiling experiments. Capture probes for the *Saccharomyces cerevisiae* gene *SWI5* (YDR146C) were designed as 50-mer standard DNA and two different LNA-modified oligonucleotides with LNA substitutions at every second or every third nucleotide position, respectively, for comparison (Table 7). To assess the sensitivity of DNA versus LNA capture

probes, hybridizations with different amounts of biotin-labelled antisense oligonucleotides in a 10-fold dilution series were performed.

#### *Design and Synthesis of the LNA Capture Probes*

To design capture probes, regions with unique mRNA sequence of the selected target 5 genes were identified. Optimized 50-mer oligonucleotide sequences with respect to  $T_m$ , self-complementarity, and secondary structure were selected. LNA modifications were incorporated to increase affinity and specificity. The biotin-labelled antisense DNA target oligonucleotide corresponds to the reverse complement sequence.

#### *Printing of the LNA Microarrays*

10 The microarrays were printed on Immobilizer™ MicroArray Slides (Exiqon, Denmark) using the Biochip One Arrayer from Packard Biochip technologies (Packard, USA). The arrays were printed with a spot volume of 2×300 pl of a 10  $\mu$ M (final concentration) capture probe dilution. Four replicas of the capture probes were printed on each slide

#### *Hybridization with Biotin-labelled Antisense Oligonucleotide*

15 The arrays were hybridized overnight using the following protocol. The desired amount of biotin-labelled oligonucleotide was combined in one tube followed by addition of 3  $\mu$ L 20×SSC (3×SSC final), 0.5  $\mu$ L 1 M HEPES, pH 7.0 (25 mM final), 25  $\mu$ g yeast tRNA (1.25  $\mu$ g/ $\mu$ L final), 0.6  $\mu$ L 10% SDS (0.3% final), and DEPC-treated water to 20  $\mu$ L final volume. The biotin-labelled target sample was filtered in a Millipore 0.22 micron spin column according  
20 to the manufacturer's instructions (Millipore, USA), and the probe was denatured by incubating the reaction at 100°C for 2 minutes. The sample was cooled at 20-25°C for 5 minutes by spinning at maximum speed in a microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was carefully placed on top of the microarray spotted on Immobilizer™ MicroArray Slide, and the hybridization mixture was applied to the array from the side. An aliquot of 30  $\mu$ L of 3×SSC  
25 was added to both ends of the hybridization chamber, and the Immobilizer™ MicroArray Slide was placed in the hybridization chamber. The chamber was sealed watertight and incubated at 65°C for 16-18 hours submerged in a water bath. After hybridization, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The LifterSlip coverslip was washed off in 2×SSC, pH 7.0 containing 0.1% SDS at room  
30 temperature for 1 minute, followed by washing of the microarrays subsequently in 1.0×SSC, pH 7.0 at room temperature for 1 minute, and then in 0.2×SSC, pH 7.0 at room temperature for 1 minute. Finally, the slides were washed for 5 seconds in 0.05×SSC, pH 7.0. The slides were then dried by centrifugation in a swinging bucket rotor at approximately 200 G for 2 minutes. To visualize the biotin containing duplexes, an aliquot of 40  $\mu$ L of the 2  $\mu$ g/ml streptavidin-Cy3



in 1×SSC+0.05 % Tween solution was applied to the slide as described for the hybridization mixture above. The slide was incubated in a humidified chamber for 1 hour at room temperature. The coverslip was washed off in 1×SSC+0.05 % Tween for 1 minute, followed by wash in 0.2×SSC+0.05 % Tween for 1 minute and then 10 seconds in MilliQ-water. The slide was dried by centrifugation in a swinging bucket rotor for 2 minutes at 200 G.

#### *Data Analysis*

Following washing and drying, the slides were scanned using a ScanArray 4000XL scanner (Perkin-Elmer Life Sciences, USA), and the array data were processed using the GenePix™ Pro 4.0 software package (Axon, USA).

#### **10 Results**

Incorporation of LNA nucleotides at every second or third nucleotide position in standard 50-mer expression array oligonucleotide capture probes results in a 2-7-fold increase in fluorescence intensity levels using an unsaturated target concentration and hybridizing under standard stringency conditions (Fig. 21). Thus, it can clearly be concluded that the LNA 15 oligonucleotides are more sensitive in expression profiling compared to DNA oligonucleotides.

Table 7. DNA and LNA-modified *SWI5* (YDR146C) oligonucleotide capture probes. LNA modifications are depicted by uppercase letters in the sequence; “mC” denotes LNA methyl cytosine.

Oligo Name	Sequence
YDR146C-50	tggaatggaacggggattatggttcgccaatgaaaactaatcaaaggt
YDR146C-50_LNA2	TgGgAaTgGaAcGgGgAtTaTgGtTtmCgmCcAaTgAaAamCtAaTcAaAgGt
YDR146C-50_LNA3	TggGaaTggAacGggGatTatGgtTtcGccAatGaaAacTaaTcaAagGt

20

#### Example 7: The Use of LNA-modified Oligonucleotides in Microarrays Provides Significantly Improved Sensitivity in Comparative Genome Hybridization (CGH).

This example demonstrates the advantages of using LNA oligonucleotide microarrays in Comparative Genome Hybridization (CGH) experiments. Capture probes for all 23 exons of the Menkes gene (*ATP7A*) were designed as 50-mer standard DNA and different LNA/DNA mixmer oligonucleotides, respectively, for comparison (Fig. 25). The C6-amino-linked capture probes were applied to Immobilizer slides and hybridized with patient DNA samples labelled with a Cy3 fluorescent dye.

#### *Design and synthesis of the LNA capture probes*

To design the capture probes, regions comprising individual exons of the Menkes gene were identified. The optimal 50-mer oligonucleotide sequences with respect to  $T_m$ , self-

complementarity, and secondary structure were selected for each exon. LNA modifications were incorporated to increase affinity and specificity. A software tool "OligoDesign", which automatically designs capture probes that are optimized for sequence specificity,  $T_m$ , self-complementarity, secondary structure, and LNA modifications was used for oligonucleotide design.

### Results

Fluorescent Cy3 labelled patient genomic DNA was hybridized to microarrays spotted with the CGH capture probes listed in Fig. 25. Compared to DNA capture probes, capture probes with LNA in every second position (LNA-2) had a significantly better capture rate of 10 non-amplified labelled genomic patient DNA as shown in Figs. 22-24. Capture probes of other lengths and/or with other LNA substitution patterns can be used similarly.

### Example 8: Expression Profiling of Stress and Toxicity in *Caenorhabditis elegans* using LNA Oligonucleotide Microarrays

15 This example demonstrates the use of the Exiqon *C. elegans* LNA tox oligoarray in gene expression profiling experiments in the nematode *Caenorhabditis elegans*. The *C. elegans* tox oligoarray monitors the expression of a selection of 110 genes relevant for general stress response and for the metabolism of toxic compounds. Two different capture probes for each of these target genes were designed and included in the LNA tox array. In addition, the *C. elegans* 20 LNA tox oligoarray contained capture probes providing control for cDNA synthesis efficiency and the developmental stage of the nematode. Capture probes for constitutively expressed genes for data set normalization were also included on the *C. elegans* LNA tox oligoarray.

### *Cultivation of C. elegans Worms*

For all cultures, the sample was divided into two, and one half of the sample was used 25 as the control, the other was used as the treated sample. Worm samples were harvested and sucrose cleaned by standard methods. For heat shock treatment, the heat shock sample was added to S-media preheated to 33°C in a 1 L flask suspended in a water bath at 33°C, the other sample was added to a 1 L flask with S-media at 25°C. Both samples were shaken at approximately 100 rpm for an hour. For Lansoprazole treatment, 0.5 mL of 10 mg/mL 30 Lansoprazole (Sigma) in DMSO was added to each 500 mL volume of S-media culture after 28 hours of growth from L1. At the same time, 0.5 mL of DMSO was added to the control. Incubation was for 24 hours. Samples were then harvested by centrifugation at 3000×g suspended in RNeasy Lysis Buffer (Qiagen) and immediately frozen in liquid nitrogen.

### *RNA Extraction*

RNA was extracted from the worm samples using the FastRNA® Kit, GREEN (Q-BIO) essentially according to the suppliers' instructions.

### *Design and Synthesis of the LNA Capture Probes*

- 5 To design the capture probes, regions with unique mRNA sequence of the selected target genes were identified. The optimal 50-mer oligonucleotide sequences with respect to  $T_m$ , self-complementarity, and secondary structure were selected. LNA modifications were incorporated to increase affinity and specificity.

### *Printing of the LNA Microarrays*

- 10 The microarrays were printed on Immobilizer™ MicroArray Slides (Exiqon, Denmark) using the Biochip One Arrayer from Packard Biochip technologies (Packard, USA). The arrays were printed with a spot volume of 2×300 pl of a 10  $\mu$ M capture probe solution. Four replicas of the capture probes were printed on each slide.

### *Synthesis of Fluorochrome Labelled First Strand cDNA from Total RNA*

- 15 15  $\mu$ g of *C. elegans* total RNA was combined with 5  $\mu$ g oligo dT primer (T20VN) in an RNase free, pre-siliconized 1.5 mL tube, and the final volume was adjusted with DEPC-water to 8  $\mu$ L. The reaction mixture was heated at +70°C for 10 minutes, quenched on ice 5 minutes, spin 20 seconds, followed by addition of 1  $\mu$ L SUPERase-In™ (20U/ $\mu$ L, Ambion, USA), 4  $\mu$ L 5×RTase buffer (Invitrogen, USA), 2  $\mu$ L 0.1 M DTT (Invitrogen, USA), 1  $\mu$ L dNTP (20mM 20 dATP, dGTP, dTTP; 0.4 mM dCTP in DEPC-water, Amersham Pharmacia Biotech, USA), and 3  $\mu$ L Cy3™-dCTP or Cy5™-dCTP (Amersham Pharmacia Biotech, USA). First strand cDNA synthesis was carried out by adding 1  $\mu$ L of Superscript™ II (Invitrogen, 200 U/mL), mixing, and incubating the reaction mixture for 1 hour at 42°C. An additional 1  $\mu$ L of Superscript™ II was added, and the cDNA synthesis reaction mixture was incubated for an additional 1 hour at 25 42°C; the reaction was stopped by heating at 70°C for 5 minutes, and quenching on ice for 2 minutes. The RNA was hydrolyzed by adding 3  $\mu$ L of 0.5 M NaOH, and incubating at 70°C for 15 minutes. The samples were neutralized by adding 3  $\mu$ L of 0.5 M HCl, and purified by adding 450  $\mu$ L 1×TE buffer, pH 7.5 to the neutralized sample and transferring the samples onto a Microcon-30 concentrator. The samples were centrifuged at 14000×g in a microcentrifuge for 30 ~8 minutes, the flow-through was discarded, and the washing step was repeated twice by refilling the filter with 450  $\mu$ l 1×TE buffer and by spinning for ~12 minutes. Centrifugation was continued until the volume was reduced to 5  $\mu$ L, and finally the labelled cDNA probe was eluted by inverting the Microcon-30 tube and spinning at 1000×g for 3 minutes.

### *Hybridization with Fluorochrome-labelled cDNA*

The arrays were hybridized overnight using the following protocol. The Cy3™ and Cy5™-labelled cDNA samples were combined in one tube followed by addition of 3 µL 20×SSC (3×SSC final), 0.5 µL 1 M HEPES, pH 7.0 (25 mM final), 25 µg yeast tRNA (1.25 µg/µL final), 0.6 µL 10% SDS (0.3% final), and DEPC-treated water to 20 µL final volume. The labelled cDNA target sample was filtered in a Millipore 0.22 micron spin column according to the manufacturer's instructions (Millipore, USA), and the probe was denatured by incubating the reaction at 100°C for 2 minutes. The sample was cooled at 20-25°C for 5 minutes by spinning at maximum speed in a microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was carefully placed on top of the microarray spotted on Immobilizer™ MicroArray Slide, and the hybridization mixture was applied to the array from the side. An aliquot of 30 µL of 3×SSC was added to both ends of the hybridization chamber, and the Immobilizer™ MicroArray Slide was placed in the hybridization chamber. The chamber was sealed watertight and incubated at 65°C for 16-18 hours submerged in a water bath. After hybridisation, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The Lifterslip coverslip was washed off in 2×SSC, pH 7.0 containing 0.1% SDS at room temperature for 1 minute, followed by washing of the microarrays subsequently in 1.0×SSC, pH 7.0 at room temperature for 1 minute, and then in 0.2×SSC, pH 7.0 at room temperature for 1 minute. Finally, the slides were washed for 5 seconds in 0.05×SSC, pH 7.0. The slides were then dried by centrifugation in a swinging bucket rotor at approximately 200 G for 2 minutes. The slide was then ready for scanning.

### *Data analysis.*

Following washing and drying, the slides were scanned using a ScanArray 4000XL scanner (Perkin-Elmer Life Sciences, USA), and the array data were processed using the GenePix™ Pro 4.0 software package (Axon, USA). The data in each image was normalized so that the ratio of means of all of the features is equal to 1.

### *Results*

Use of LNA-modified oligonucleotide capture probes in the *C. elegans* LNA tox oligoarray clearly allows the identification of distinct expression profiles for *C. elegans* genes relevant for general stress response and for the metabolism of toxic compounds.

Table 12. Expression profiling using LNA Oligonucleotide Microarrays. Log2 transformed fold of changes for five selected genes in the two expression profiling experiments.

Protein name

<u>(clone name)</u>	<u>Heat shock</u>	<u>Lansoprazole</u>
HSP70 (F44E5.4/5)	4.11	nd
CYP37A (F01D5.9)	nd	0.98
Ubiquitin (M7.1)	0.16	-0.12
Histone 1Q (C01B10.5)	-1.49	nd
HSP90 (C47E8.5)	nd	-1.17

5 Table 13. LNA-modified oligonucleotide capture probes. LNA modifications are depicted by uppercase letters in the sequence; "mC" denotes LNA methyl cytosine.

<u>Oligo Name</u>	<u>Sequence</u>
CEABC_C34G6.4_u293_LNA3	TgcmCatTgcAcgGgcActTgtTcgAtcTccTtcTgtTttActTttGgaTg
CEABC_C34G6.4_u375_LNA3	TcaTtcTagGatTgcmCagAtgGttAtgAtamCtcAtgTcgGagAgaAagGa
CEABC_F57C12.4_u15_LNA3	mCcaAtgTgtTttAatTggTgtTaaTgtmCttGatGacmCtgmCatAatmCatAt
CEABC_F57C12.4_u480_LNA3	mCacAagAtcmCtgTgtTgtTctmCcgGaamCaaTgaAaaTgaActTagAtcmCa
CEABC_F57C12.5_u111_LNA3	TacTgtTtcTcgAcaAagGttGtgTagmCcgAgtTtgAcamCtcmCgaAgaAa
CEABC_F57C12.5_u444_LNA3	TgaActTggAtcmCctTctTgtmCatTtaGcgAtgAtcAaaTttGggAagmCg
CEABC_K08E7.9_d8_LNA3	TcaTtaAttTgtTgtAgcTttmCttTctmCgaTttTgtmCacGatmCttTccmCc
CEABC_K08E7.9_u51_LNA3	AggGtgmCctActAcaAacTgamCccAaaAgcAgaTgamCcgAgaAgaAatAa
CEABC_Y39D8C.1_u37_LNA3	AttGaaAgcGacGcgGaaAgtGccAtgTatTtcTaaTttTgtTttmCttTa
CEABC_Y39D8C.1_u422_LNA3	TtgTcaGcaTatmCaaGagTagAlaTggAagTggAtamCacTctGctAatmCc
CEADH_H24K24.3a_d3_LNA3	mCacmCttAttGcgTtcAatTttTgtTtcmCacmCtamCtamCtamCgaAtamCgtTg
CEADH_H24K24.3a_u50_LNA3	TcamCaaGggAgaGagTctGcgGtcGgtGctGgcGttmCgaGaaAatAtaAc
CEAPEX_R09B3.1_u191_LNA3	mCatGcaTccmCgamCgaGaaGaaGtamCtcAttTtgGagTtaTctGgcGaaTt
CEAPEX_R09B3.1_u37_LNA3	GacmCatGctmCcgGtcGtcAtgmCaaAtcGacTtcTaaAttGctTctGatTa
CEAPO_C35D10.9_u15_LNA3	TtgmCatGctGttAaaAccTatmCgtGtamCaaTatTgcmCtgTatAttmCccmCt
CEAPO_C35D10.9_u609_LNA3	TggmCacAgcTtaAtaAcaAatTggAaaGtcGagGatTagTcgGtgTtgAa
CEAPO_C48D1.2_u176_LNA3	GacAcamCgcAaaGgaTatGgaTgtTgtTgaGctGctGacTgaAgtmCaaTa
CEAPO_C48D1.2_u23_LNA3	AgcAcgAaamCtcTgcmCgtmCtaAaaTtcActmCgtGatTcaTtgmCccAatTg
CEAPO_F20C5.1_u453_LNA3	AtgGtcAtamCtcTaaAatGggmCagAacTtcAacmCaaAtcAttmCtcGtcAg
CEAPO_F20C5.1_u96_LNA3	AacmCcgAgcTtgmCcgmCaaAgtGcaAgaAaaTtaTagAacGaaTgaAacAg
CEATPase_B0365.3_u31_LNA3	GgaTggGtcGagmCgtGagAccTacTaaAgaAcaGctTgtGaaTctTt mCaamCgtTctmCgaTtcmCtamCggAcaAgaAtgGacmCtaTgcmCaamCagAaa
CEATPase_B0365.3_u386_LNA3	Ga
CEATPase_C17H12.14_u356_LNA3	TgcTcgTtaTccAgcTatTttGaaGggActTgtmCatGcaAggActTctTc
CEATPase_C17H12.14_u89_LNA3	mCcgTttAgaGctTatTgcTaamCcaGatTgtmCccAcaAgtmCagAacAgcTc
CEATPase_F55F3.3_u215_LNA3	TgamCggAcgmCtamCtamCccAtaTgtAttTgtTccAtcTtamCcaGcaAccAa

CEATPase\_F55F3.3\_u275\_LNA3 AgcTatTtcAttmCgamCaaGgaAcaTctmCggAaaAgtmCaaGtamCatmCccGg  
 CEATPase\_Y49A3A.2\_u103\_LNA3 AaaTtcAagGatmCcaGttGccGatGgtGaaGccAagAttmCgcAagGatTa  
 CEATPase\_Y49A3A.2\_u272\_LNA3 mCgaTcgTttmCtgmCccAttmCtamCaaGacTgtmCggTatGctmCaaGaaTatGa  
 CECALR\_Y38A10A.5\_u238\_LNA3 TcaGgaAagAacTttGacAacAttAtcAtcAccGacTctGttGagGagGc  
 CECALR\_Y38A10A.5\_u296\_LNA3 TgaActmCtamCtcTtaTgaAagmCtgGggAgcmCatmCggAttmCgaTttGtgGc  
 CECAT\_Y54G11A.5b\_u137\_LNA3 GaamCttTgcAggGccGctmCggGgaAtgTcaTgaTttmCatTatTaaGggAa  
 CECAT\_Y54G11A.5b\_u189\_LNA3 GtcAatTctGggAgaAggTgtTggAtamCcgGggmCtcGggAgaGaaTgtGc  
 CECC\_C03D6.3\_u275\_LNA3 AtgTaaAgaAggAatGctTccmCgaAtgGatTggAtaTttAttTgtmCcaGa  
 CECC\_C03D6.3\_u430\_LNA3 GgamCcgAaaTttGtgmCagmCatGtcGgamCacGaaAttGatGgtmCtcAttTt  
 CECC\_C07G2.3\_d9\_LNA3 mCagAcamCgaAggTtamCgaTagAtaAccAtcTctmCaaAgtmCtaTcgAccTc  
 CECC\_C07G2.3\_u44\_LNA3 mCgamCgaTgtGcgTgtTccTgamCgaTgaAagAatGggAtaTtaAgaAaamCc  
 CECC\_Y46G5A.2\_u331\_LNA3 TtgTgcTccAtcGctGctmCcgmCttAcaGacTtgAcaAcgmCtcAccTttGc  
 CECC\_Y46G5A.2\_u385\_LNA3 AatGagmCggTgtTgcmCgtGtgAcgTcamCttmCgtmCacAgtGttGctmCtamCt  
 CECOA\_C29F3.1\_u316\_LNA3 AaaTtgAcamCcaAtcAaaTctGtcTcaTctmCctGagGacmCgtmCaamCttmCg  
 CECOA\_C29F3.1\_u392\_LNA3 AatmCttTgtGtamCggAgaTggGgcAaaAggmCagmCaaGaaAgtAaamCcaAg  
 CECOA\_F08A8.4\_u1094\_LNA3 AggAcaAggGgcActActGgcAcaGgcTttGatTatTgcAgtGagAtaTt  
 CECOA\_F08A8.4\_u1260\_LNA3 TtaAtgGagGtgAcaAtgGgtTccTtgGatTcgAtaAatTccGagTgcmCc  
 CECOA\_F59F4.1\_u109\_LNA3 GctmCttmCtcmCagTggGctmCaaAatAgtmCaamCtcAacAgaTcgGaaGttmCt  
 CECOA\_F59F4.1\_u424\_LNA3 AaaGctTcgAgaTggmCacGttmCgtmCtgTatmCtcGtgAagAacTtaTtgmCa  
 CECOA\_Y25C1A.13\_u115\_LNA3 GatTcgmmCtgAacTttAtcAagAcgTggAatAigAgcmCagmCtcmCtgTcgAc  
 CECOA\_Y25C1A.13\_u451\_LNA3 GatmCttAtcAccGcgTgcGatAttmCgaGtaGctTcamCagGatGcgAttTt  
 CECOL\_C27H5.5\_u493\_LNA3 GgaAagGaaGgaTccAttmCtcAgcTctGcamCttmCcamCcaTcaGagmCcaTg  
 CECOL\_C27H5.5\_u680\_LNA3 TggAtamCaaGgaGggAicTggmCagTggTggAicTggAagTggTggAtaTg  
 CECOQ\_ZC395.2\_u199\_LNA3 TtgAaaGaamCtcmCttGccGacGatmCctGaaAcamCacAaaGaaTtgmCtgAa  
 CECOQ\_ZC395.2\_u400\_LNA3 AtgTggGatGagGagAaaGaamCatTtaGatAcaAtgGaaAgaTtaGctGc  
 CECRYZ\_F39B2.3\_u171\_LNA3 AggmCtgAgcTctTggActTtgGcaTcaAcaTtgTctmCatTctTgaAggAa  
 CECRYZ\_F39B2.3\_u222\_LNA3 TtaTggTtamCagAagGagmCtgTttAcgGtgTagmCatTggGaaTgtmCttmCc  
 CECyclin\_R02F2.1a\_u24\_LNA3 mCacTtcAacmCaamCtcmCgtGttAatmCaaGcaAgcmCgcmCacmCatmCtaAtg  
 Ag  
 CECyclin\_R02F2.1a\_u312\_LNA3 TctmCatTgcTcgTcgAggmCtamCcaAcaAacActGgcAatAccmCaaTtaAt  
 CECyclin\_ZC168.4\_u203\_LNA3 TaaGaaAgtmCatTgaGgaTgcTgtmCgcTttGctmCgcmCgaAgtmCtcGtaTa  
 CECyclin\_ZC168.4\_u273\_LNA3 AagTtcAtcmCtgTtgAcgGaaTcgAggmCggAgaAtgmCtgTatmCggTcaTt  
 CECYP\_B0213.15\_u133\_LNA3 AcaGgaAatAtgAttTtgGatTtcGatTttGaaTcgGttGgtGctGccmCc  
 CECYP\_B0213.15\_u202\_LNA3 GctGagmCtgTatTtgGctAgtGaaAtgTgtGttTttGatActTtaAatGa  
 CECYP\_B0304.3\_u38\_LNA3 AcgAggTttGgaTcamCaaTcaGaaTtcTgtGaaAtaAgcGttTttTggGa  
 CECYP\_B0304.3\_u89\_LNA3 AgtTctmCggTctAacAgtGtcTccmCgtTgaAtaTtcTtgTaaAatmCacAc  
 CECYP\_C03G6.14\_u706\_LNA3 AtgAccActmCaaAatActGctAaaAgaTttGcaGcgGcaGaaGccGttAa  
 CECYP\_C03G6.14\_u768\_LNA3 TtgAtaTggmCtgTacmCtgTatGgtTttTgaGgamCgtTttTtaGgaGtcGa  
 CECYP\_C03G6.15\_d9\_LNA3 AttTatTcaTtcAtcmCatGtaAacTgtAtaTttTgaAttTgtGttGtaAa  
 CECYP\_C03G6.15\_u148\_LNA3 GccAaaGcaGaaTtgTatTtgAtcTtcGgtAacmCttmCtcmCttmCgcTacAa  
 CECYP\_C06B3.3\_u102\_LNA3 AttTtgAatmCttmCtgGgaAaaTgcmCatmCcamCtcGagAaamCcgTtcmCgtTt  
 CECYP\_C06B3.3\_u474\_LNA3 mCtaAcgGagGatmCtcGccAatTatmCttTgaGagAcaAaamCtgAaamCtcmCt  
 CECYP\_C12D5.7\_u399\_LNA3 AtcTagTccmCaaTgaAtcTccmCacAtgmCtgTtamCtcGtgAtgTtcAacTc  
 CECYP\_C12D5.7\_u65\_LNA3 TttTgcTttmCatmCgcAaaAgcTcaAgaTtamCacAtgTcaGgtmCaaGccAa

CECYP\_C45H4.17\_u27\_LNA3  
 CECYP\_C45H4.17\_u598\_LNA3  
 CECYP\_C45H4.2\_u110\_LNA3  
 CECYP\_C45H4.2\_u429\_LNA3  
 CECYP\_C49C8.4\_u363\_LNA3  
 CECYP\_C49C8.4\_u883\_LNA3  
 CECYP\_C49G7.8\_d6\_LNA3  
 CECYP\_C49G7.8\_u795\_LNA3  
 CECYP\_F01D5.9\_u374\_LNA3  
 CECYP\_F01D5.9\_u46\_LNA3  
 CECYP\_F08F3.7\_u25\_LNA3  
 CECYP\_F08F3.7\_u401\_LNA3  
 CECYP\_F14F7.2\_u397\_LNA3  
 CECYP\_F14F7.2\_u68\_LNA3  
 CECYP\_F42A9.5\_u435\_LNA3  
 CECYP\_F42A9.5\_u55\_LNA3  
 CECYP\_K07C6.3\_u3\_LNA3  
 CECYP\_K07C6.3\_u354\_LNA3  
 CECYP\_K07C6.4\_u118\_LNA3

CECYP\_K07C6.4\_u87\_LNA3  
 CECYP\_K07C6.5\_u7\_LNA3  
 CECYP\_K07C6.5\_u99\_LNA3  
 CECYP\_K09A11.3\_u362\_LNA3  
 CECYP\_K09A11.3\_u48\_LNA3  
 CECYP\_K09A11.4\_u238\_LNA3  
 CECYP\_K09A11.4\_u68\_LNA3  
 CECYP\_K09D9.2\_u151\_LNA3  
 CECYP\_K09D9.2\_u866\_LNA3  
 CECYP\_T10B9.10\_u410\_LNA3  
 CECYP\_T10B9.10\_u56\_LNA3  
 CECYP\_T10B9.7\_u102\_LNA3  
 CECYP\_T10B9.7\_u267\_LNA3  
 CECYP\_T19B10.1\_u100\_LNA3  
 CECYP\_T19B10.1\_u319\_LNA3  
 CECYP\_Y49C4A.9\_u121\_LNA3  
 CECYP\_Y49C4A.9\_u413\_LNA3  
 CECYP\_ZK177.5\_u394\_LNA3  
 CECYP\_ZK177.5\_u445\_LNA3  
 CEDAO\_C47A10.5\_d9\_LNA3  
 CEDAO\_C47A10.5\_u269\_LNA3  
 CEDC\_C01A2.3\_u373\_LNA3  
 CEDC\_C01A2.3\_u96\_LNA4  
 CEDC\_C34F6.1\_u301\_LNA3

mCcg mCgam CttTaa AgaGaa Gatm CatAaa TttGca TtgTttTgtTgtAt  
 mCga GggTga TtcGga GacTttm CagTaa Tgtm CcaActTtcAaa TgtTgtmCa  
 TagAtam CaaGatAcaTccm CtcAaa AgaAggm CctAccGtcAatGgcm CaaAg  
 TcaAcgm Cgtm CtaTaa AtgAatm CacAacGagGtaTcaAcaTtcTccm CccTg  
 Atgm CtgAtgTtgAaa Tgtm CtgGctAccGtaTtc mCaaAagAtam CtgTaaTc  
 AtgAatm CcaTggm CttGgam Catm Ctc mCcgTttTtcAagGgaTatAaaAatGt  
 Atgm Caam CgaAttAgTgaaAaaTtcAtcm CtgGaaTaaAaaAtaAttm CtaAa  
 AtcGctAcgAcaAtcTttm CcgAtgm CctTcgAagTttm CgaAagm CttTctm Ct  
 GagGtcGgtGgaGgaGgaAgtGgaAatTgam Cggm CaaAatm CctGccm CaaGg  
 mCccTctTtgGgaTttm Ccam CtcAagTttActGttm Cggm Cagm CagTgaTatAa  
 GagTtgGttm Ccam CagAatGctTagGacGttTaaAttm Cgtm CacAaam CttTt  
 mCaaTatGgtTccm CatTttAgcAacTcaTatGaam CacAgaAgaTgtm CctTg  
 GaaAaaGgcGtcGacAttTtaTgtGacAcgTggAcam Cttm CacTatGacAa  
 TaaTtgAatTacGggTctTttGlam CatAttAatTttAgTAtam CttTgtGa  
 AtaTcaAtgm Caam CtaTtaAtgAatm CacAacGtcTgtm CcaAtcTtcTccm Cg  
 GgaGtgActAtgAaaGcaAagAgTtcm CgaTtgAaam CtgAaaGacAgam Ca  
 Aatm CttTaaTgaTaaTttAtgGgaTctGtaTttm CtcTttm CtgTcaAtaAa  
 AtgAgcm Ccam CaaAtgTaaAagGatAcgAgaTtgAttm CggGaam CagTcaTg  
 Atcm Ctg mCgaTatGacAttAagm Ccam CatGgtTctGaam CctTcaAcaGaaGa  
 mCtgAacm Cttm Caam CagAagAtaAacTtc mCgtAtaGcg mCtgGaaAaam Ctc mC  
 t  
 AttTaaAggAatTcam Cagm CtcAaaAaaTaaTaam Ctm CcgGttm CagAgaTt  
 AatTtgAgcm CacAtgGcaAgTtatm Caam CagAggAgam CaaTgcm CgtAcaGt  
 Tgam CatTctActTaaAggGaaGaaAatAccAacTggTacm CctTgtAttTg  
 Tcam Ccam CaaAgcm CatAcaTatGcgAgcTagTtc mCtcAggm Ctg mCttAaam Cc  
 TtcGacAaaActAttTtgGaaAgaAcaAtcm CcaTtcAgTgtGgcAaam Cg  
 TctGacAacAaaGccAtam CacGtg mCcgActAatTccAcaAtcAgcTagAa  
 TtgGcaAaaGcaGaaTtgTatTtaAtcTttGgaAacm Ctc mCttm Cttm CgcTa  
 TgaAtcTttm CaaActTatm CacTccTttTaaTacTacm CgtTccTgtTtgGa  
 AttGagAttGtaTccAttGgcGtcTctTgtTcam CaaTcgAaaAtgTctm Ca  
 AacTgcTacTatTgcGccAtcAagTgtGctGctm CaaActTaaAtcm CagGt  
 TtgAgam CagGaaAtaAgam CtaGaaTtc mCttTgaAacTggTggGaaGtg mCt  
 AagAtgTcaAagAatTcaAgcm CagAacGatGgtm Ccam CcgAcgAgcm CatTa  
 AttGaam CcaActm CtgAaaTatAatGacAcaAaam CcaTgtm CtgGaaGtgGt  
 GgcAatGtgAcaAtaTctm CcaAtgGttm Cttm CacAgcAatm Catm CacGtgTt  
 mCtaTtcAatm CgaTatTttAtcAcam CcaTccAgTgtGgam Cctm CcaTcaTt  
 GtcTcaGagAtgTgtAaaTttActTccm Ctg mCaaTttGttTcam CgcAacTa  
 Ttc mCgaAtgTttm CcaAttGggActGaaGttTcaAgaGtcAccm CagAaaAa  
 Gatm CcaGcaTctTccAagm CttAcaTtc mCctm CgtGctTgtAtcAagGaaAc  
 TttGaaAacm CtgTttTatTaaAatAgaTaaTtgAttAgTctGtam Cg  
 Atam CgtTgcActGcaTccGgcTatGagGgaGccAaaAatm CttAggGgaGt  
 Gcam Cttm CcaTtcAtcTctGcaGctActAtgGctTtgGtgAcaAaaGttGg  
 mCcgTccAaaAgaAtgm CcaTctm CacAagTctTgaAatm CttAtaAagGtaGt  
 GagGgaTcaAcaGtaAccTcgTgcGgtAttGacAagGgaTgtm CcgGaaGg

CEDC\_C34F6.1\_u450\_LNA3  
 CEDC\_F33D11.3\_u126\_LNA3  
 CEDC\_F33D11.3\_u14\_LNA3

CEDC\_F46E10.2\_u392\_LNA3  
 CEDC\_F46E10.2\_u54\_LNA3  
 CEDC\_F56G4.2\_u382\_LNA3  
 CEDC\_F56G4.2\_u82\_LNA3  
 CEDC\_M162.2\_u103\_LNA3  
 CEDC\_M162.2\_u480\_LNA3  
 CEDC\_R10E4.11\_u274\_LNA3  
 CEDC\_R10E4.11\_u397\_LNA3  
 CEDC\_T04C9.1\_u321\_LNA3  
 CEDC\_T04C9.1\_u64\_LNA3

CEDC\_W02A2.3\_u32\_LNA3

CEDC\_W02A2.3\_u374\_LNA3  
 CEDC\_W05G11.3\_u153\_LNA3  
 CEDC\_W05G11.3\_u51\_LNA3  
 CEDC\_ZK863.5\_u256\_LNA3  
 CEDC\_ZK863.5\_u324\_LNA3  
 CEEPHX\_Y55B1BR.4\_u161\_LNA3  
 CEEPHX\_Y55B1BR.4\_u93\_LNA3  
 CEER\_18S\_u388\_LNA3  
 CEER\_18S\_u82\_LNA3  
 CEER\_26S\_u342\_LNA3  
 CEER\_26S\_u38\_LNA3  
 CEFOXO\_R13H8.1b\_u331\_LNA3  
 CEFOXO\_R13H8.1b\_u393\_LNA3  
 CEGAPDH\_K10B3.7\_u21\_LNA3  
 CEGAPDH\_K10B3.7\_u727\_LNA3  
 CEGBA\_F11E6.1a\_u232\_LNA3  
 CEGBA\_F11E6.1a\_u451\_LNA3  
 CEGLU\_C02A12.1\_u264\_LNA3  
 CEGLU\_C02A12.1\_u55\_LNA3  
 CEGLU\_C46F11.2\_u271\_LNA3  
 CEGLU\_C46F11.2\_u45\_LNA3  
 CEGLU\_F26E4.12\_u109\_LNA3  
 CEGLU\_F26E4.12\_u480\_LNA3  
 CEGLU\_R07B1.4\_u166\_LNA3  
 CEGLU\_R07B1.4\_u38\_LNA3  
 CEGLU\_T09A12.2\_u220\_LNA3  
 CEGLU\_T09A12.2\_u335\_LNA3

GatGgtTctTcgAtcGcaAacAaaAcaGatGtgmCtcmCatTtamCatAcgGa  
 AtgGagAaaAtgGatmCtgAtgGagTtgmCagGaaGtgAtgGagmCtcmCagGa  
 TgaAtcTccAtaAatTatTcaAtgTttmCcaAatAttTaaTttAtcAatTg  
 GctmCaamCacGgtAggAtcmCtaTggAacmCgtmCggAggAgcAggmCctmCggA  
 g  
 mCgtGacAacmCtcTtaTttAttTctGtaAaamCtgAttmCgcmCaaActTttGt  
 GaaGctTtcAaamCcaAatGagTtcmCttmCccGgaAtcmCcaAagAatAccAa  
 AcaAtgAaaAgaGagGatGgaAagGaaAtcGaaGtcTctGttmCttGacGa  
 GatGagGtamCatAacTttGtgTgcAgtTatAggmCcaTctAcaGtamCctGc  
 TtcmCatmCatmCacTaamCcgAttGtcmCtgAcaTtgAtgGccAaamCcaGggAa  
 TcamCatTatmCgaAcaAgtActAgtAagmCatGctGtgAtgGagTgcmCgcTa  
 mCacGgaGatmCacGacAtcAaaGcgGatTgcTtaGagTgtGgaAacmCgtmCt  
 ActAtcTacGtgGcamCgtTggActmCatmCatmCgaTggGaamCgamCgtAtaAg  
 TctmCtgGccAgtTcamCttTgtGatmCaaTctmCagAttmCgtmCcamCacAagAt  
 mCtamCttmCogmCaaGaaGgcmCcgTcgTttmCtaAtcGatmCgaAcaTctmCacA  
 c  
 AtgGatGatmCgamCccActTgcmCacTgamCccAcaAtcmCcgmCacTcamCtam  
 Cc  
 AagAcgGagAggmCtgGagAgaAcgGtamCcgAtgGagAgcmCagGaamCtgAt  
 mCcamCccAggAggAggGatAcaAgaGaaGaaAgtAcaGatTctmCcaActAa  
 AgtTtcAcamCttmCttTttGccGttTtgGttmCccGttAtcAatmCcaTtgAt  
 mCttTtaTatTctmCatmCaaTttGttTccTacTtgGtcAgtTgaGgaTcgTt  
 TtcGgcAcaAatGgaGcaAaaGtaTcgTggTtaTtgTgaTgcGatTatTc  
 mCtamCtaTgaAtgAgtTcamCtgGacTcaTttAtcAacTcgAgtmCaaAagmCc  
 GttGgcGaaTctTcgGgtTcgTatAacTtcTtaGagGgaTaaGcgGtgTt  
 GaamCtgAttmCgaGaaGagTggGgamCtgTcgmCttmCgaGgtTtaAcgActTc  
 TgtTatTgcGaaAgtAatmCctGctTagTatGagAggAacAgtGggTtcAa  
 TgcAtamCgamCttGgtmCtcTtgGtcAagGtgTtgTatTcaGtaGagmCagTc  
 TgtGctmCagAatmCcamCttmCttmCgaAatmCcaAttGtgmCcaAgtActAacTt  
 TtaAgamCggAacmCaaTtgmCtcmCacmCacmCatmCatAccAgtTgaAcaGt  
 AcaTtgmCtamCcaAggmCctAagmCcgmCttmCaaAttmCtcTaaGtcTgaAatGa  
 GttGagTccAccGgaGtcTtcAccAccAtcGagAagGccAatGctmCacTt  
 AgtAaaTtcmCttmCcamCgtGgaTctActmCgtGtgTtcAcaAagAtcGagGg  
 GgtmCcaAtaAtgGgaGacTggTtcmCgcGcaGaaAgtTatGcaGatGatAt  
 AgaAaamCttmCgtTggAccmCtgmCtaAggAgaAgtAttTcaAgtTtcTgaGc  
 GagmCacmCcgAagmCtcAagmCcaTatTtgGaaAcaAgamCcaTacTctTcaAa  
 GttAccmCtcTacAaaTctmCgcTtcAatmCcaAtgTtgTtcGcaGtcAccAa  
 mCcgAagAgtTcgTtamCtaTgcGagGagGtgTgaAgcmCggAatAatTttTt  
 AagTtcTtgGttGgamCgcGatGggAaaAttAtcAagAgaTttGgamCcaAc  
 AcgAttTcaAcgTcaAaaAtgmCtaAtgGtgAtgAcgTgtmCacTttmCggAt  
 AccTggGttGatGttTttGcgGctGaaAgtTtcTccAagmCtcAttGatTa  
 GaaGtamCgtmCtcmCcaAagAaaAgtTacmCccAgtTtaAggmCatTgcAcaAt  
 GcgcmCcaGatAtgTatTcaAagAtcGagGtaAatGgtmCagAacActmCatmCc  
 AatmCtamCagGgaAaaAggAttTcgAgtTgcmCgcGttTccAtgmCaaTcaAt



CEGLU\_T28A11.11\_u299\_LNA3  
 CEGLU\_T28A11.11\_u54\_LNA3  
 CEGPD\_B0035.5\_u256\_LNA3  
 CEGPD\_B0035.5\_u478\_LNA3  
 CEHSP\_C09B8.6\_d8\_LNA3  
 CEHSP\_C09B8.6\_u286\_LNA3  
 CEHSP\_C12C8.1\_u127\_LNA3  
 CEHSP\_C12C8.1\_u1531\_LNA3  
 CEHSP\_C47E8.5\_u310\_LNA3  
 CEHSP\_C47E8.5\_u361\_LNA3  
 CEHSP\_F26D10.3\_u276\_LNA3  
 CEHSP\_F26D10.3\_u397\_LNA3  
 CEHSP\_F43D9.4\_u169\_LNA3  
  
 CEHSP\_F43D9.4\_u275\_LNA3  
 CEHSP\_F44E5.4/5\_u123\_LNA3  
 CEHSP\_F44E5.4/5\_u380\_LNA3  
 CEHSP\_F52E1.7\_u175\_LNA3  
 CEHSP\_F52E1.7\_u448\_LNA3  
 CEHSP\_F54D5.8\_u252\_LNA3  
 CEHSP\_F54D5.8\_u318\_LNA3  
 CEHUS\_H26D21.1\_u117\_LNA3  
 CEHUS\_H26D21.1\_u478\_LNA3  
 CEMRE\_ZC302.1\_u169\_LNA3  
 CEMRE\_ZC302.1\_u292\_LNA3  
 CEMTL\_T08G5.10\_d127\_LNA3  
 CEMTL\_T08G5.10\_u45\_LNA3  
 CENAP\_D2096.8\_u356\_LNA3  
 CENAP\_D2096.8\_u70\_LNA3  
 CEPAL\_F56D12.5\_u241\_LNA3  
 CEPAL\_F56D12.5\_u301\_LNA3  
 CEPDI\_C07A12.4\_u28\_LNA3  
 CEPDI\_C07A12.4\_u433\_LNA3  
  
 CEPDI\_C14B1.1\_u119\_LNA3  
 CEPDI\_C14B1.1\_u358\_LNA3  
 CEPKG\_T03F1.3\_d9\_LNA3  
 CEPKG\_T03F1.3\_u424\_LNA3  
 CEPON\_E01A2.7\_u223\_LNA3  
 CEPON\_E01A2.7\_u79\_LNA3  
  
 CEPPGB\_F13D12.6\_u44\_LNA3  
 CEPPGB\_F13D12.6\_u440\_LNA3  
 CEPPS\_T14G10.1\_d2\_LNA3

AgaTggmCaaAgaAgaCatAacTgaAacTctTccmCggGgaGctActAc  
 TgaAtaAacGggmCcgAacTaaAtcmCatTcgTcaGtgGaaAtgGgaAacAa  
 GtcmCgtmCttmCctGatGctTatGaamCgcmCtaTttmCtcGaaGtaTtcAtgGg  
 TgtGgaAaaGctmCtcAacGagAagAaaGcaGaaGttmCgtAtamCaaTtcAa  
 AtaTcgmCcgcmCctGctTccTcamCcaAccmCgaAtaAcgmCaamCaaAaamCttTa  
 AagAgcmCcamCtcAicAagGatGaaAgtGatGgaAagActmCttmCgtmCtcAg  
 mCaaGatAttTtaAcaAaaAtgmCatmCaamCaaGaaGccmCaaTcaGgtTccGg  
 mCttGggmCatTctGtamCggGatGctGtcAttActGtgmCctGcaTatTtAa  
 AagAagmCatmCtcGaaAicAacmCcaGacmCacGctAicAtgAagAcamCttmCg  
 AtgAaaGctmCaaGctmCttmCgtGatTccTctActAtgGgaTacAtgGccGc  
 TtaAgcAgamCcaTtgAggAgaAgaAgcTcaAggAtaAgaTcaGccmCagAa  
 mCgtmCttTccAagGatGacAttGaamCgcAtgGtcAacGaaGctGagAaaTa  
 GtcGacTtgGctmCacAtcmCacAccGtcAicAacAagGaaGgamCagAtgAc  
 mCaaTctTgaGggAcamCgtTctmCacmCatTgaGggAcamCcamCgaGgtmCaa  
 Ga  
 TcamCtaAaaTgcAccAatmCtgGacAatmCttmCtgmCttmCtgmCtgGatGcgcmCt  
 TcaTgaAgcTaaAcaAttmCgaAaaGgaAgaTggTgaAcaAagGgaAagTg  
 AagTatAacmCttmCcaAcaGggGtcmCgtmCcaGaamCaaAicAagTccGaaTt  
 TttAacmCatGgcmCgcAgaTtcTtcGatGacGtcGacTttGatmCgcmCacAt  
 GcgTcgAaaAgaTctmCccTgaAgtmCtgmCatTgamCtgGccTtgAtaTtaTg  
 AcaTagTctTcgTcaTcaAggAtaAgcmCacAccmCgaAatTcaAgcGagAg  
 TcgmCcaAcamCtcGgamCacGtgmCcaAaaTgaAtaTcaTctmCaaAicGaaTg  
 GtcGaaGttAgaAatmCcaGaaGccGatAttGttTctmCatmCaaAttmCcaAt  
 ActActmCgtGgaAgaTccAatAaaGttGttTcaAcgmCgamCaaAicGatTc  
 GgcAgtGaaGatGaaGtgGcaAatTctGatGaaGaaAtgGgaAagAgtAt  
 TtgTcaAagAccAgaAgaAaaAatTatGggAatmCgcGatAaaAttmCaaGg  
 GatGcaAgtGtgmCcaActGcgAatGtgmCtcAggmCtgmCtcAttAatTtgAa  
 GacGatAtgTtcGatTtcmCcaGgaGagGacGgtGatGatGtgTcaGacTt  
 GacGatAtgTtcGatTtcmCcaGgaGagGacGgtGatGatGtgTcaGacTt  
 GagGtcGtcGtaAtcmCacAagGctmCcaAgaAagmCaaGtgmCtcGacAttTc  
 GatActTttGgcAagmCtcGttmCcaAicAagAagGagGtcAtcmCcaGatmCg  
 GatGagGagGgamCacAccGagmCtcTaaAtcmCacAttmCcaAtamCagTtcAa  
 mCttAtgTccGaaGatAtcmCcaGagGatTggGacAagAacmCcaGtcAagAt  
 TacmCccAgtmCgamCtaTgaTggAgamCagAaamCctmCgaGaaGttmCgaAgaA  
 t  
 mCtcGtcGccTccAacTicAacGaaAttGccmCttGatGaaAccAagActGt  
 TtcTatTgtTtaTtcmCttGccmCaaTagTgtAttTgtAttTatTctTtcTc  
 mCaaAtcmCatmCtcmCcaGtgGatTtcGtcAttGctGacAagTtcGccGagGa  
 GttTctGatTcgAcamCttTatGgamCcaTctmCaaGttmCtgmCgaGttTctTt  
 GggAaamCaaAtgAttGttGgtAcaGtaGccmCgcmCctGctAttmCacTgtGa  
 mCgaGcamCatmCatmCcaAicGttmCctGttmCaamCaaGgcmCttmCtaAicGttA  
 g  
 TgaTgaGagmCccAgtAacmCaaTtaTttGaamCcgTcaGgaTgtGcgTaaGg  
 mCgtmCtaAicGaaGaaGggGatmCgtGggmCaaTcaTaamCtaAttAacmCttmCa

CEPPS\_T14G10.1\_u240\_LNA3  
 CEPRDX\_R07E5.2\_u405\_LNA3  
 CEPRDX\_R07E5.2\_u42\_LNA3  
 CEPYC\_D2023.2\_u256\_LNA3  
 CEPYC\_D2023.2\_u427\_LNA3  
 CERAD\_F10G7.4\_u169\_LNA3

CERAD\_F10G7.4\_u267\_LNA3  
 CERAD\_F32A11.2\_u250\_LNA3  
 CERAD\_F32A11.2\_u380\_LNA3  
 CERAD\_T04H1.4\_u274\_LNA3  
 CERAD\_T04H1.4\_u375\_LNA3  
 CERAD\_W06D4.6\_u325\_LNA3  
 CERAD\_W06D4.6\_u34\_LNA3  
 CERAD\_Y116A8C.13\_u289\_LNA3  
 CERAD\_Y116A8C.13\_u59\_LNA3  
 CERAD\_Y39A1A.23\_u221\_LNA3  
 CERAD\_Y39A1A.23\_u276\_LNA3  
 CERAD\_Y41C4A.14\_u509\_LNA3  
 CERAD\_Y41C4A.14\_u731\_LNA3  
 CERAD\_Y43C5A.6\_u131\_LNA3  
 CERAD\_Y43C5A.6\_u429\_LNA3  
 CERFC\_F31E3.3\_u128\_LNA3  
 CERFC\_F31E3.3\_u55\_LNA3  
 CERPL\_K11H12.2\_d1\_LNA3

CERPL\_K11H12.2\_u172\_LNA3  
 CERT\_F36A4.7\_u1396\_LNA3  
 CERT\_F36A4.7\_u2302\_LNA3  
 CERT\_F36A4.7\_u289\_LNA3  
 CERT\_F36A4.7\_u2919\_LNA3  
 CERT\_F36A4.7\_u4269\_LNA3  
 CERT\_F36A4.7\_u5485\_LNA3  
 CESLC\_F52F12.1a\_u249\_LNA3

CESLC\_F52F12.1a\_u76\_LNA3  
 CESLC\_K11G9.5\_u400\_LNA3  
 CESLC\_K11G9.5\_u462\_LNA3  
 CESLC\_Y32F6B.1\_u179\_LNA3  
 CESLC\_Y32F6B.1\_u280\_LNA3  
 CESLC\_Y37A1C.1a\_u104\_LNA3  
 CESLC\_Y37A1C.1a\_u404\_LNA3  
 CESLC\_Y70G10A.3\_u383\_LNA3  
 CESLC\_Y70G10A.3\_u46\_LNA3

mCaaTggmCtcmCagGtcTttmCtgmCtcTtcAtaTacTtcmCatTccGagTtgmCt  
 GttmCtcTtgGagmCtgAagTtgTcgcmCgtGctmCgtGtgAttmCtcActTctmCt  
 TcgcmCtamCcaGcaAggAatActTcaAcaAggTcaAcaAgtGatmCacAcaGa  
 AagGaaAttGtaActmCgcmCcaAgaGctmCtcmCcaGgtGtcmCgtGgamCatAt  
 TtgActGgaTtgGagAttGcgGaaGaaGttGatGttGaaAtcGagAgtGg  
 GccAagTctmCaaGcaAtaAgtGttGatmCaaTcaGagmCcaTacGgaGagAt  
 AtaTtgAgamCttmCggGacAagmCggActTctmCatmCtgTcamCagmCaamCtgm  
 Cc  
 GatmCcgcmCagAgaAtcGagTatTtcmCtcTcgAgamCccAtgGatAtcAacTg  
 TccGttAagAagmCtcActGgaAaaAcamCacGgcTcgAacGaaAttGgaAt  
 AatTtgGatGagAgaAaaGtgGaaGgaAtgGctAtcGttTtgGcaGatAt  
 GtgmCtgGtcAaaAaaTgcTtgmCttmCgtTgcTtaTtcGcaTtgmCacTcgcmCa  
 mCttmCgaGaamCtcTtcAagTtgGaaTcaAcaGtgGcaTcgGatAcamCatGa  
 GtgmCctTctGaaGccGaaGaaAacGacGatTagTtaAatGttTccAagTt  
 GatAaaAtcGatAgcGacGacGatGagGaaGccGatGatGagGagmCtcGa  
 GcaGgtGgaTacGgaTgtGgaGctGacTttTgcGttTtaTcaAgaAtcTc  
 TccmCgtAgaAgtAgaAatGctAgaAgaAccTgaAcaAgaAgaTcaAgaAa  
 TgcAagAtgTcaGtaTtgAaamCaaTtcmCtgTagAgamCccmCcgAagAaaAt  
 AgtmCtcGtaTccGggAatGttTcaGccTgtGaaAatGctTgtTgaAgamCg  
 mCttmCaaAacmCgtmCgcTttTaaGgaTacAggAacGtgGcamCgcTtcmCgaGg  
 mCagAttGtamCctTcgAaaAggAaaAggAgaGaaTcgcmCgtmCgcAaaAatGg  
 TgaTggmCttTgaTtaTtcGagmCagGagmCaaTgaTgtmCcgAgaGtcGttAt  
 mCaaTgamCgaGaaTatTggAgtAatGggGaaActGgtTgcGacTtgmCgaAa  
 TtgGaaAacAatmCtcmCtcGacTttmCtgmCtcActmCttmCgtGaaActAtcmCa  
 TctTgtTatTttAttTtgTttTggGctTgtTccGaaAatGaaAtgGttGt  
 mCaaTggAtcAccAagmCcaGttmCacAagmCacmCgtGagmCaaAgaGgamCtc  
 Ac  
 mCttTgtGatGtgAtgActGcgAagGgamCacTtgAtgGctAttAcgAgamCa  
 GagmCcaGctActmCagAtgAcamCtcAacAcgTtcmCatTatGcaGgaGttC  
 TacActmCcaTccTcgcmCcgAcaTcaAatmCcaAcaTctmCcamCgcGgaTtcTc  
 AtgGagAagAtgGttTggAtgGaaTgtGggTtgAgaAtcAgaAtaTgcmCg  
 AacmCggGatAccGtgTcgAacGtcAcaTgaAagAtgGcgAtaTaaTcgTc  
 GagGagAttAaamCgcAtgTcaGtgGctmCatGtcGagTttmCcaGaaGtcTa  
 AgaTatTgcmCtcTacTtaTcaTggGccTgaTggmCttTgtmCtgmCcgGtaTt  
 GaaTctmCaamCcamCttmCtgGaamCccmCatAcamCcaAtgGatAgaAgamCgg  
 Ag  
 GttGttmCttTttTccGtgAtcTttTcaTgtTtaTgtmCtgAacGtgGcaGg  
 GacTcgTtgGtgTctTgcTagGatGtcTtgGgtTcaTtcmCtcAatmCgtTg  
 GtamCtgGgcTcgAggGctGaaActAatmCgaAgaAgaAacTccAgaAgaTa  
 GgaTcaTgcTctGttTacGacActGatGagTtaAgaGtcAgamCtgmCacGt  
 mCgaTggTtcTtcTcgTctAtcAtaTcgGggTagTtgmCcgAagTgtTgaAa  
 mCaaAtcGaamCtgGtaTaaAggAggAccGacGgaGacGaaTttGaamCgaGa  
 AttmCgaTcaAagAacTctGgcTctmCggmCgtTaamCtgGacAttTgtTcgTc  
 mCtcmCccGagmCagGcgAttAttmCacGctAgtTatGctmCaaAtgTgaTctGt

CESOD\_C15F1.7\_u435\_LNA3  
 CESOD\_C15F1.7\_u9\_LNA3  
 CESOD\_F10D11.1\_u326\_LNA3  
 CESOD\_F10D11.1\_u477\_LNA3  
 CESULT\_EEED8.2\_u316\_LNA3  
 CESULT\_EEED8.2\_u82\_LNA3  
 CESULT\_Y113G7A.11\_u252\_LNA3  
 CESULT\_Y113G7A.11\_u96\_LNA3  
 CESULT\_Y67A10A.4\_u108\_LNA3  
 CESULT\_Y67A10A.4\_u327\_LNA3  
 CETOPO\_K12D12.1\_u398\_LNA3  
 CETOPO\_K12D12.1\_u449\_LNA3  
 CETOPO\_M01E5.5b\_u256\_LNA3  
 CETOPO\_M01E5.5b\_u429\_LNA3  
 CEUbi\_F25B5.4\_u186\_LNA3  
 CEUbi\_F25B5.4\_u2\_LNA3  
 CEUbi\_F29B9.6\_u145\_LNA3  
 CEUbi\_F29B9.6\_u230\_LNA3  
 CEUbi\_M7.1\_u239\_LNA3  
 CEUbi\_M7.1\_u53\_LNA3  
 CEUGT\_F39G3.1\_u40\_LNA3  
 CEUGT\_F39G3.1\_u466\_LNA3  
 CEUGT\_M88.1\_u480\_LNA3  
 CEUGT\_M88.1\_u72\_LNA3  
 YAL009W\_u145\_LNA3  
 YAL009W\_u341\_LNA3  
 YAL059W\_u262\_LNA3  
 YAL059W\_u51\_LNA3  
 YER109C\_u109\_LNA3  
 YER109C\_u436\_LNA3  
 YHR152W\_u128\_LNA3  
  
 YHR152W\_u510\_LNA3  
 YKL130C\_u211\_LNA3  
 YKL130C\_u85\_LNA3  
 YKL178C\_u199\_LNA3  
 YKL178C\_u367\_LNA3  
 YLR443W\_u179\_LNA3  
 YLR443W\_u86\_LNA3  
 YOR092W\_u251\_LNA3  
 YOR092W\_u82\_LNA3  
 YPL263C\_u132\_LNA3  
 YPL263C\_u257\_LNA3

mCcgGtamCtaTctGgaTcamCacAgaAgtmCcgAaaAtgAccAggmCagTtaTt  
 mCccAgtGacTacmCtgAatmCgcGtcTctGaaTctmCcamCacAatTccTacTa  
 GgaGttGctmCacmCgcAatTaaGagmCgamCttmCggAtcTctGgaTaaTctTc  
 AaaTtgAggAaaAgcTicAcgAggmCggTctmCcaAagGaaAcgTcaAagAa  
 mCaaTcgTacmCatGaaAgaAgTggAagmCcamCgtGcaAgaGaaGaaAtcmCa  
 AagAagAttmCctGacmCagAgaGacTcamCgtGctTacmCcaAgaAgcAtcTa  
 AgcAttGgtGgaAatAcgAaaTggmCatGggAagAgaAacmCccTctmCaaTt  
 mCtgGttAcgGtaGtgTatGgtmCccTgtmCctmCtcAgaAtgmCaaAtaTgtmCg  
 TctAcgTcgAtgGaaAagmCcgAttTaamCaaTcaAagmCcaAcaAcgmCagTt  
 GgaAagGtgmCcaAaaAgTgamCagmCaaTtgGagGatmCttAttmCatTgcmCa  
 AgaTgaTgaTgaAgTtccTgcAaaGaaGccTgcTccAgcGaaGaaAgcTg  
 AaaAccTcgTacTggAaaAggAgcTgcGaaAgcGgaAgTatmCgaTttGt  
 GagAagGccmCagAagAagTacGacAgamCtgAagGagmCagTtgAaaAagTt  
 TtcTgtmCatAcaAtcGtgmCtaAtcGgcAggTgtmCgaTccTttGtaAccAt  
 AagmCttmCggAcamCcaTtgAgaAtgTcaAagmCcaAaaTccAggAtaAggAg  
 AatmCgaAccmCatmCaaTtcActmCgtTatTccTccTcgAtcTccGttmCaaGt  
 mCtgAacmCatmCcaAatAttGaaGatmCcaGctmCagGctGaaGccTatmCagAt  
 mCgtGtgmCttAtcTctTctGgaTgaAaamCaaGgaTtgGaaGccGtcAatmCt  
 mCggAagmCatmCtgmCctTgamCatTctmCcgTtcGcaGtgGtcGccGgcTctG  
 AaaGtamCgcTatGtgAggAggmCtaAcamCcaTtcAtaTaaGaamCgcAgcmCa  
 TgtTgcmCgtAgaAgaGagActAaaActAagAacGatTgaTtgAagGtcTg  
 TacAatTctTtgmCagGaaGcaAtaTccGccGgaGtcmCccmCttAicActAt  
 mCtcAcgGagGttAtaAttmCtaTgcAggAggmCaaTttmCtgmCtgGagTcmCa  
 AccGttTcaTgaGagmCtgTaaTcaGgtGttGttTctGtaAaaAgTgtGaa  
 GtgGatGtgAaaTtaGtcmCtcAacmCccAgaGcaTttAgTgcaGagAttAg  
 GcaGttTaaTgtGaaGctAgTaaAgTacaGtcTacGtgGgamCgaGaaAt  
 AttGccAagTccAttTctmCgtGccAagTacAttmCaaAatAcaAgaAagGc  
 AgamCtcmCtamCaaAtaGatTcgGtgTccTgcmCagAcgAtgTtgAagAatAg  
 TtgAagTttGggAatAttGgtAtgGttGaaGacmCaaGgamCcgGatTacGa  
 GagGcgmCaaGtaGgcAatGatTcaAgaAgTAgTaaaGgcAatmCgtAacAc  
 TgaGcamCaaAgTaaGatGttmCggAaaGaaAaaGaaAgtmCaaTccTatGa  
 mCaaGtgAccAatmCagmCacGcamCggmCttmCcaTccTcaAgamCtgAtaTtam  
 Cc  
 AttAaaTgcGcaGatGagGacGgaAcgAatAtcGgaGaaActGatAatAt  
 GatGgtAagmCtgAgcGccTtgGacGaaGaaTttGatGttGtcGctActAa  
 TacGtcAcgmCaaGgamCagAgcTttGacGacGaaAtaTcamCttGgaGgaTt  
 TctmCccTgtGtaGgtAcamCcaAtaTcamCaaGcgmCatTtcTatGtcGacTa  
 TgcTaamCacmCagTttAgamCcaTggAaaTccmCacmCgcAaaTatAagmCaaTg  
 GcaGgamCatAagAttmCcgGtcAagmCaamCgamCagTgaAgaAagTatGcaAa  
 mCcgTctAgTgaaAgcGggAtgGctAaaTtgGgaAaamCgamCaaGatGttAt  
 GatGctTcaAtaTccTttGatGgtmCgtTagTttAccAttTttGgtGtcTt  
 iCatTtgAgTatGtgAagAccGttGgtGggAaaGaaGagAtcAggTg  
 GtcTtgGctAccAcamCccAaaAccGttmCgaAacTttAagAgcAttmCtamCt

**Example 9. Performance Analysis of LNA Oligonucleotide Capture Probes Designed to Detect Ratios of Splice Variants in mRNA Pools.**

***Oligonucleotide Design for Mmicroarrays.***

The methods for designing exon-specific internal oligonucleotide capture probes are 5 described above.

***Design of the LNA-modified Capture Probes***

For the internal LNA-modified oligonucleotide capture probes, every third DNA nucleotide was substituted with an LNA nucleotide. The probes designed to capture the junction of the recombinant splice variants were designed with LNA modifications in a block of 10 five consecutive LNAs nucleotides, two on the 5' side of the splice junction and three on the 3' side of the splice junction. All capture probes are shown in Table 14.

Table 14. Internal, exon-specific and merged, exon-exon junction specific oligonucleotide capture probes used in this example.

<u>Capture probes</u>	<u>Sequence (LNA=uppercase; DNA lowercase letters)</u>
gene78.01a_50_LNA3	mCctGaaAgtAgaTttGttAttTccGaaAcgmCctTctmCccGttmCttAagTc
gene78.01b_50_LNA3	mCatAtamCcamCaaAtaGtcmCctmCaaAaaTcamCaaGaaAacTcamCaamCacTg
gene78.03a_50_LNA3	GatTtgmCagmCggTggTaaAaaGtaTgaAaamCgtGgtAatTaaAagGtcTc
gene78.03b_50_LNA3	mCcaAtgAaaActAatmCaaAggTaaAcgTggAtcmCcaTggmCaaTtcmCcgGg
gene78.m01INS3_50_block	caacactgcccagaggttcaatcGATmCmCgatgatcctaataagggcgccc
gene78.mINS303_50_block	gtccagtatcgtccatcatAGTATcgataaatatgtgaaggaaatgcctg
gene78.m01INS4_50_block	caacactgcccagaggttcaatcGATGTgtgataggatcagtggtcaggg
gene78.mINS403_50_block	gaaggcgaaggagactgctAATATcgataaatatgtgaaggaaatgcctg

15

***Printing and Coupling of the Splice Isoform-specific Microarrays***

The splice variant capture probes were synthesized with a 5' anthraquinone (AQ)-modification, followed by a hexaethyleneglycol-2 (HEG2) linker. The capture probes were first diluted to a 20  $\mu$ M final concentration in 100 mM Na-phosphate buffer pH 7.0, and spotted on 20 the Immobilizer polymer microarray slides (Exiqon, Denmark) using the Biochip Arrayer One (Packard Biochip Technologies, USA) with a spot volume of 2x 300 pl and 300  $\mu$ m between the spots. The capture probes were immobilized onto the microarray slide by UV irradiation in a Stratalinker with 2300  $\mu$ joules (Stratagene, USA). Non-immobilized capture probe oligonucleotides were removed from the slides by washing the slides two times 15 minutes in 25 1xSSC. After washing, the slides were dried by centrifugation at 1000xg for 2 minutes, and stored in a slide box until microarray hybridization.

### *Construction of Splice Variant Vectors*

The recombinant splice variant constructs were cloned into the Triamp18 vector (Ambion, USA). The constructs were sequenced to confirm their construction. The plasmid 5 clones were transformed into *E. coli* XL10-Gold (Stratagene, USA).

#### *Triamp18/SWIS Vector Construct*

Genomic DNA was prepared from a wild type standard laboratory strain of *Saccharomyces cerevisiae* using the Nucleon MiY DNA extraction kit (Amersham Biosciences, USA) according to the supplier's instructions. Amplification of the partial yeast gene was performed using standard PCR using yeast genomic DNA as template. In the first step of amplification, a forward primer containing a restriction enzyme site and a reverse primer containing a universal linker sequence were used. In this step, 20 bp was added to the 3'-end of the amplicon, next to the stop codon. In the second step of amplification, the reverse primer was exchanged with a nested primer containing a poly-T<sub>20</sub> tail and a restriction enzyme site. The SWIS amplicon contains 730 bp of the SWIS ORF plus 20 bp universal linker sequence and a poly-A<sub>20</sub> tail.

The PCR primers used were;

YDR146C-For-EcoRI: acgtgaattcaaatacagacaatgaaggagatga

YDR146C-Rev-Uni: gatccccgggaattgccatgttacctttgattagtttcattggc

20 Uni-polyT-BamHI: acgtggatcctttttttttttttttgatccccgggaattgccatg.

The PCR amplicon was cut with the restriction enzymes, *EcoRI* + *BamHI*. The DNA fragment was ligated into the pTRIamp18 vector (Ambion, USA) using the Quick Ligation Kit (New England Biolabs, USA) according to the supplier's instructions and transformed into *E. coli* DH-5 $\alpha$  by standard methods.

#### 25 *Construction of the Recombinant Splice Variant #1 (Triamp18/swi5-rubisco)*

The *Arabidopsis thaliana* Rubisco small subunit ssu2b gene fragment (gi17064721) was amplified from genomic DNA by primers named DJ305 5'-ACTATGATGGACGATACTGGAC-3' and DJ306 5'-ATTGGATCGATCCGATGATCCTAATGAAGGC-3', containing *ClaI* restriction site linkers. The purified PCR fragment was digested with *ClaI* and then cloned into the swi5 (gI:7839148) vector at the unique *ClaI* site (atcgat) giving each insert a flanking sequence from the original yeast SWIS insert (named exon01 and exon 03, Fig. 19). The product was inserted in the reverse orientation, so that the insert sequence is:

atcgatCCGATGATCCTAATGAAGGCGCCCGGGTACTCCTTCTTGCATTCTTCAACTTC  
 CTTCAACACTTGAGCGGAGTCGGTGCATCCGAACAATGGAAGCTTCCACATTGTCC  
 AGTATCGTCCATCATAGTatcgat

5 Nucleotide sequence analysis revealed a difference between the sequence of *A. thaliana* rubisco expected from the GenBank database and that obtained from all sequenced constructs and PCR products. Position 30 in the Rubisco insert is "C" rather than the expected "A". This SNP was probably created by PCR. None of the oligonucleotide capture probes used in the example cover this region. Rubisco sequence in genbank is TCCTAATGAAGGCGCCA. The sequence  
 10 obtained

from the plasmid construct is TCCTAATGAAGGCGCCC.

#### *Construction of the Recombinant Splice variant # 2 (Triampl8/swi5-lea)*

The *Arabidopsis thaliana* Lea gene (gi1526423) was amplified from genomic DNA with primers named DJ307 5'-GGAATTATCGATGTGTGATAGGATCAGTGTTTCAG-3' and  
 15 DJ308 5'-AATTGGATCGATATTAGCAGTCTCCTTCGCC-3' including the *Cla*I linker sites as above. The PCR fragment was digested with *Cla*I cloned into the yeast SW15 IVT construct as above at the unique *Cla*I site. The fragment was inserted in the forward orientation, resulting in the following insert sequence:

atcgatGTGTGATAGGTTTCAGTGTTTCAGGGCTGTCCAAGGAACGTATGAGCATGCGAG  
 20 AGACGCTGTAGTTGGAAAAACCCACGAAGCGGCTGAGTCTACCAAAGAAGGAGCT  
 CAGATAGCTTCAGAGAAAGCGGTTGGAGCAAAGGACGCAACCGTCGAGAAAGCTA  
 AGGAAACCGCTGATTATACTGCGGAGAAGGTGGGTGAGTATAAAGACTATACGGT  
 TGATAAAGCTAAAGAGGCTAAGGACACAACCTGCAGAGAAGGCGAAGGAGACTGCT  
 AATatcgat

25

#### *Preparation of Target*

##### *In vitro* RNA Preparation from Splice Variant Vectors

*In vitro* RNA from the splice variants were made using the MEGAscript™ high yield transcription kit according to the manufacturer's instructions (Ambion, USA). The yield of  
 30 IVT RNA was quantified at a Nanodrop spectrophotometer (Nanodrop Technologies, USA).

##### Isolation of Total RNA from *C. elegans*

*C. elegans* wild-type strain (Bristol-N2) was maintained on nematode growth medium (NG) plates seeded with *Escherichia coli* strain OP50 at 20 °C, and the mixed stages of the  
 35 nematode were prepared as described in Hope, I. A. (ed.) " *C. elegans* – A Practical Approach

“, Oxford University Press 1999. The samples were immediately flash frozen in liquid N<sub>2</sub> and stored at – 80 °C until RNA isolation.

A 100 µl aliquot of packed *C. elegans* worms from a mixed stage population was homogenized using the FastPrep Bio101 from Kem-En-Tec for 1 minute, speed 6 followed by 5 isolation of total RNA from the extracts using the FastPrep Bio101 kit (Kem-En-Tec) according to the manufacturer's instructions. The eluted total RNA was ethanol precipitated for 24 hours at –20°C by addition of 2.5 volumes of 96% EtOH and 0.1 volume of 3M Na-acetate, pH 5.2 (Ambion, USA), followed by centrifugation of the total RNA sample for 30 minutes at 13200 rpm. The total RNA pellet was air-dried and redissolved in 10 µl of diethylpyrocarbonate 10 (DEPC)-treated water (Ambion, USA) and stored at – 80°C.

#### Fluorochrome-labelling of the Target

The following fluorochrome-labelled cDNA targets were synthesized to test the performance of 'merged' probes that span exon borders. Synthetic RNAs corresponding to the splice variant #1 (exon01-INS3-exon03 (1-INS3-3) and splice variant #2 (exon01-INS4-exon03 15 (1-INS3-3) were spiked into 10µg of *C. elegans* reference total RNA sample in two different ratios. The first target pool (KU007) contained 10 ng of splice variant #1 (1-INS3-3) transcript and 2 ng of variant #2 (1-INS4-3) transcript, a ratio of 5:1. The second target pool (KU008) contained 2 ng variant #1 (1-INS3-3) transcript and 10 ng of splice variant #2 (1-INS4-3) transcript, a ratio of 1:5. Both mRNA pools were combined in separate labeling reactions with 20 5 µg anchored oligo(dT<sub>20</sub>) primer and DEPC-treated water to a final volume of 8 µl. The mixture was heated at 70°C for 10 minutes, quenched on ice for 5 minutes, followed by addition of 20 units of Supersasin RNase inhibitor (Ambion, USA), 1 µl dNTP solution (10mM each dATP, dGTP, dTTP and 0.4 mM dCTP, and 3 µl Cy5-dCTP, Amersham Biosciences, USA), 4 µl 5 x RTase buffer (Invitrogen), 2µl 0.1 mM DTT (Invitrogen), 400 units of Superscript II 25 reverse transcriptase (Invitrogen, USA) and DEPC-treated water to 20 µl final volume. Background hybridization to merged capture probes was monitored in both hybridizations using the other fuor channel with 10µg of *C. elegans* reference RNA alone labeled with Cy3-dCTP, according to the labeling method described above for the splice variant spikes. All four cDNA syntheses were carried out at 42°C for 2 hours, and the reaction was stopped by incubation at 30 70°C for 5 minutes, followed by incubation on ice for 5 minutes.

Unincorporated dNTPs were removed by gel filtration using MicroSpin S-400 HR columns as described below. The column was pre-spun for 1 minute at 1500 xg in a 1.5 ml tube, and the column was placed in a new 1.5 ml tube. The cDNA sample was slowly to the top center of the resin, spun 1500-xg for 2 minutes, and the eluate was collected. The RNA was

hydrolyzed by adding 3  $\mu$ l of 0.5 M NaOH, mixing, and incubating at 70 °C for 15 minutes. The samples were neutralized by adding 3  $\mu$ l of 0.5 M HCl and mixing, followed by addition of 450  $\mu$ l 1xTE, pH 7.5 to the neutralized sample and transfer onto a Microcon-30 concentrator (prior to use, 500  $\mu$ l 1xTE was spun through the column to remove residual glycerol). The 5 samples were centrifuged at 14000-xg in a microcentrifuge for 12 minutes. Spinning was continued until volume was reduced to 5  $\mu$ l. The labelled cDNA probes were eluted by inverting the Microcon-30 tube and spinning at 1000-xg for 3 minutes.

#### *Microarray Hybridization*

The fluorochrome-labelled cDNA samples, respectively, were combined (the two 10 different ratios separately). The following were added: 3.75  $\mu$ l 20x SSC (3x SSC final, which was passed through a 0.22  $\mu$ filter prior to use to remove particulates) yeast tRNA (1  $\mu$ g/ $\mu$ l final) 0.625  $\mu$ l 1 M HEPES, pH 7.0 (25 mM final, which was passed through 0.22  $\mu$ filter prior to use to remove particulates) 0.75  $\mu$ l 10 % SDS (0.3 % final) and DEPC-water to 25  $\mu$ l final volume. The labelled cDNA target samples were filtered in Millipore 0.22  $\mu$  filter spin column 15 (Ultrafree-MC, Millipore, USA) according to the manufacturer's instructions, followed by incubation of the reaction mixture at 100 °C for 2-5 minutes. The cDNA probes were cooled at room temp for 2-5 minutes by spinning at maximum speed in a microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was carefully placed on top of the microarray spotted on Immobilizer™ MicroArray Slide, and the hybridization mixture was applied to the array from 20 the side. An aliquot of 30  $\mu$ L of 3xSSC was added to both ends of the hybridization chamber, and the Immobilizer™ MicroArray Slide was placed in the hybridization chamber (DieTech, USA). The chamber was sealed watertight and incubated at 65°C for 16-18 hours submerged in a water bath. After hybridization, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The slides were washed sequentially by 25 plunging gently in 2 x SSC/0.1% SDS at room temperature until the cover slip falls off into the washing solution, then in 1x SSC pH 7.0 (150 mM NaCl, 15 mM Sodium Citrate) at room temperature for 1 minute, then in 0.2 x SSC, pH 7.0 (30 mM NaCl, 3 mM Sodium Citrate) at room temperature for 1 minute, and finally in 0.05 x SSC (7.5 mM NaCl, 0.75 mM Sodium Citrate) for 5 seconds, followed by drying of the slides by spinning at 1000 xg for 2 minutes. 30 The slides were stored in a slide box in the dark until scanning.

#### *Microarray data analysis*

The splice variant microarray was scanned in a ScanArray 4000XL confocal laser scanner (Packard Instruments, USA). The hybridization data were analysed using the GenePix Pro 4.01 microarray analysis software (Axon, USA). Only the Cy5 (650 nm) data were



examined as both hybridizations produced comparable, and acceptably low, signal from the *C. elegans* reference RNA alone (Cy3 channel).

#### Normalization

Data was normalized so that it could be compared between hybridizations. Both 5 hybridizations contained the same amount of RNA from synthetic exons 01 and exons 03 (10+2 ng), so signal from the capture probes designed to internal regions of these exons is expected to be equal. The ratio of raw Cy5 signal between the two different labeled cDNA target pools, designated as KU007 and KU008 hybridizations, for each probe corresponding to either of these exons was calculated, that is for each probe  $i$  we calculated the ratio  $\text{probe}_i\text{KU007}/\text{probe}_i\text{KU008}$ . The average of all of these ratios was used as the normalization ratio.

#### Expectations of normalized data.

To reflect the proportions of RNA spiked into the hybridization, the ratio of signal in hybridization KU007/KU008 should be 5 for probes designed to exon junctions of the INS3 splice variant #1 and 0.2 for probes corresponding to 1-INS4 splice variant #2. Data was  $\log_2$  transformed:  $\log_2(5)=+2.32$ ,  $\log_2(0.2)=-2.32$ . The merged probe corresponding to the exon 01-exon 03 border desirably produces a consistently low value that is desirably independent of which transcript was more abundant, i.e.,  $\log_2(\text{ratio})=0$ .

#### Array results

Results are summarized in Table 15. 50-mer capture probes containing LNA in a block 20 spanning exon-exon junctions were consistent in producing the expected ratios.

Table 15. LNA 50-mer block probes are most consistent in producing overall data closest to expected ratios.

Capture probe	Expected ratio ( $\log_2$ )	Observed ratio ( $\log_2$ ) with merged LNA block capture probes
gene78.m0103	0.00	-0.24
gene78.m01INS3	2.32	2.93
gene78.m01INS4	-2.32	-2.39
gene78.mINS303	2.32	3.11
gene78.mINS403	-2.32	-0.86

Example 10: Improved Signal-to-noise Ratios using LNA Oligonucleotide Capture Probes Combined with cDNA Target Fragmentation with the *E. coli* Uracil-DNA Glycosylase.

### Capture Probe Design

The capture probes were designed to a 602 nucleotide 3'-region of the Yeast (*S. cerevisiae*) 70 kDa heat shock protein (SSA4) gene. The 602-base pair sequence is shown in Table 16. For the LNA-spiked oligonucleotide capture probes, every third DNA nucleotide was substituted with a LNA nucleotide. All capture probes are shown in Table 17.

Table 16. Six hundred and two (602) base sequence stretch of *S. cerevisiae* *ssa4* gene. The underline indicates the position of the capture probes. First underline is equal to capture probe YER103W-554, second underline is equal to capture probe YER103W-492 and so forth.

<u>ggcgaaggacaaaggacaaaagacaaacatctactgggtaaatbtgagtgagcggatattccacccgctccaagaggegtaccac</u>
<u>aaattgaagttacatttgatctcgatgc aaatggatctctgaacgtatctgccttgaaaaaggtactggtaaatctaacagat</u>
<u>tacaattactaacgat aagggaagattatcgagggaagat atcgataaaatgggtgctgaggcagaaaaagttcaaggccgaagat</u>
<u>gaacaaagaagctcaacgtgttc aagct aagaatcagctagaatcgtacgcgtttactttgaaaaattctgtgagcgaaaaataact</u>
<u>tcaaggagaaggtgggtgaagaggatgccaggaaattggaagccgcgcgcccaagatgctataaatgggttagatgcttcgc aagc</u>
<u>ggcctccaccgaggaatacaaggaaaggcaaaaggaaactagaaggtgttgcaaacccattatgagt aatatttacggagctgca</u>
<u>ggcgggtgcccaggagcaggccagttccgggtgctggagcaggccccactggagcaccagacaacggcccaacgggtgaagagg</u>
<u>ttgattag</u>

10

Table 17. Capture probes for the SSA4 tile array.

Oligo Name	Sequence
YER103W-1-DNA	gccccactggagcaccagacaacggcccaacggtgaagaggttgattag
YER103W-38-DNA	gccccaggagcaggccagttccgggtgctggagcaggccccactggagc
YER103W-73-DNA	ccattatgagtaaaatttacggagctgcagggtggtgccccaggagcaggc
YER103W-92-DNA	ctagaaggtgtgcaaaccattatgagtaaaatttacggagctgcagg
YER103W-127-DNA	cctccaccgaggaatacaaggaaaggcaaaaggaaactagaaggtgtgca
YER103W-200-DNA	ggtgaagaggatgccaggaaattggaagccgcgcccaagatgctataaa
YER103W-245-DNA	acttgaaaaattctgtgagcgaaaataactcaaggagaaggtgggtga
YER103W-272-DNA	aagaatcagctagaatctgacgcgtttactttgaaaaattctgtgagcga
YER103W-336-DNA	aatggtgctgaggcagaaaaagttcaaggccgaagatgaacaagaagctc
YER103W-393-DNA	taacaagattacaattactaacgataagggaagattatcgaaggaagata
YER103W-447-DNA	cgatgcaaatggtattctgaacgtatctgccgttgaaaaaggtactgga
YER103W-492-DNA	acccgctccaagaggcgtaccacaaattgaagttacattgatalcgtg
YER103W-554-DNA	ggtgaaaggacaaggacaaaagacaacaatctactgggtaaatttgagt
YER103W-1-LNA1	GccmCcamCtgGagmCacmCagAcaAcgGccmCaamCggTtgAagAggTtgAttAg
YER103W-38-LNA1	GccmCcaGgaGcaGgcmCcaGttmCcgGgtGctGgaGcaGgcmCccActGgaGc
YER103W-73-LNA1	mCcaTtaTgaGtaAatTttAcgGagmCtgmCagGtgGtgmCccmCagGagmCagGc

YER103W-92-LNA1	mCtaGaaGgtGttGcaAacmCccAttAtgAgtAaaTttTacGgaGctGcaGg
YER103W-127-LNA1	mCctmCcamCcgAggAatAcaAggAaaGgcAaaAggAacTagAagGtgTtgmCa
YER103W-200-LNA1	GgtGaaGagGatGccAggAaaTtgGaaGccGccGccmCaaGatGctAtaAa
YER103W-245-LNA1	ActTtgAaaAatTctGtgAgcGaaAatAacTtcAagGagAagGtgGgtGa
YER103W-272-LNA1	AagAatmCagmCtaGaaTcgTacGcgTttActTtgAaaAatTctGtgAgcGa
YER103W-336-LNA1	AatGgtTgcTgaGgcAgaAaaGttmCaaGgcmCgaAgaTgaAcaAgaAgcTc
YER103W-393-LNA1	TaamCaaGatTacAatTacTaamCgaTaaGggAagAttAtcGaaGgaAgaTa
YER103W-447-LNA1	mCgaTgcAaaTggTatTctGaamCgtAtcTgcmCgtTgaAaaAggTacTggTa
YER103W-492-LNA1	AccmCgcTccAagAggmCgtAccAcaAatTgaAgtTacAttTgaTatmCgaTg
YER103W-554-LNA1	GgtGaaAggAcaAggAcaAaaGacAacAatmCtamCtgGgtAaaTttGagTt

#### Control capture probes

YFL039C-50	acaagaatacgcacgaaagtgggtccatctatcggtcaccacaagtgtttct
YFL039C-50_LNA3	AcaAgaAtamCgamCgaAagTggTccAtcTatmCgtTcamCcamCaaGtgTttmCt
YDR146C-50	tggaatggaacggggattatggttcgccaatgaaaactaatcaaaggt
YDR146C-50_LNA3	TggGaaTggAacGggGatTatGgtTtcGccAatGaaAacTaaTcaAagGt

#### *Printing and Coupling of the Yeast SSA4 Tile Microarrays*

The SSA4 capture probes were synthesized with a 5' anthraquinone (AQ)-modification, 5 followed by a hexaethyleneglycol-2 (HEG2) linker. The capture probes (Table 17) were first diluted to a 20  $\mu$ M final concentration in 100 mM Na-phosphate buffer pH 7.0, and spotted on the Immobilizer microarray slides (Exiqon, Denmark) using the Biochip Arrayer One (Packard Biochip Technologies) with a spot volume of 2x 300 pl and 400  $\mu$ m between the spots. The capture probes were immobilized onto the microarray slide by UV irradiation in a Stratalinker 10 with 2300  $\mu$ joules (Stratagene, USA). Non-immobilized capture probe oligonucleotides were removed from the slides by washing the slides two times 15 minutes in 1xSSC. After washing, the slides were dried by centrifugation at 1000xg for 2 minutes, and stored in a slide box until microarray hybridization.

#### *Yeast Cultures*

15 *Saccharomyces cerevisiae* wild-type (BY4741, MATa; his3 $\Delta$ 1; leu2 $\Delta$ 0; met15 $\Delta$ 0; ura3 $\Delta$ 0) and  $\Delta$ ssa4 (MATa; his3 $\Delta$ 1; leu2 $\Delta$ 0; met15 $\Delta$ 0; ura3 $\Delta$ 0; YER103w::kanMX4) mutant strains (EUROSCARF) were grown in YPD at 30°C until the A<sub>600</sub> density of the cultures reached 0.8. Half of the cultures were collected by centrifugation and resuspended in one volume of 40°C preheated YPD. Incubation was continued for an additional 30 minutes at 20 30°C or 40°C for the standard and heat-shocked cultures, respectively. Cells were harvested by centrifugation and stored at -80°C.

### RNA Extraction

Total RNA was extracted using the FastRNA Kit-RED (BIO 101) according to suppliers' instructions. The quantity and quality of the RNA preparations were examined by standard spectrophotometry on a NanoDrop ND-1000 and by gel electrophoresis. Only high quality RNA preparations were used for microarray analyses.

### Fluorochrome-labelling of the Target

A total of seven cDNA assay mixtures were produced. Each with ten (10) µg total RNA from wt-*ssa4* wt and combined with 5 µg anchored oligo(dT<sub>20</sub>) primer and DEPC-treated water to a final volume of 8 µl. The mixtures were heated at 70°C for 10 minutes, quenched on ice for 5 minutes, followed by addition of 20 units of Supersasin RNase inhibitor (Ambion, USA), 3 µl Cy3-dCTP (Amersham Biosciences), 10mM final concentration of dATP and dGTP, 4 µl 5 x RTase buffer (Invitrogen), 2µl 0.1 mM DTT (Invitrogen), 400 units of Superscript II reverse transcriptase (Invitrogen, USA), dUTP and dTTP accordingly to Table 18, and DEPC-treated water to 20 µl final volume. A parallel set-up was made with 10 µg total RNA from *Δssa4* for target cDNA labelling with Cy5-dCTP. All cDNA syntheses were carried out at 42°C for 2 hours, and the reaction was stopped by incubation at 70°C for 5 minutes, followed by incubation on ice for 5 minutes. Each cDNA pool (except the unfragmented control pool) was incubated at 37°C for 2 hours with 2 units of Uracil-DNA Glycosylase (UDG, New England Biolabs, USA) and by addition of 2.4 µl (1x final concentration in the reaction mixture) of UDG reaction buffer. The enzyme was heat-inactivated at 95°C for 10 minutes. Unincorporated dNTPs were removed by gel filtration using MicroSpin S-400 HR columns as described in Example 9.

Table 18. dUTP and dTTP ratios in cDNA target labelling.

Assay #	Final conc. dUTP	Final conc. dTTP
1	0.5	0
2	0.25	0.25
3	0.125	0.375
4	0.05	0.45
5	0.025	0.475
6	0.0125	0.4875
7	0	0.5

### *Gel electrophoresis of the cDNA target pools*

0.5 µl of each of the seven fragmented cDNA pools were analysed on a 2% agarose-gel. The data show that the cDNA is fragmented linearly with respect to the concentration of dUTP used in the synthesis. Fig. 38 shows the gel electrophoresis of fragmented cDNA from the yeast wild-type strain.

### *Comparative Hybridization of the SSA4 Tile Array with Fluorochrome-labelled wild-type and Asa4 cDNA Target Pools and Post-hybridization Washes*

The fluorochrome-labelled cDNA samples, respectively, were combined (the different 10 UDG-fragmented samples separately). The following were added: 3.75 µl 20x SSC (3x SSC final, pass through 0.22 µ filter prior to use to remove particulates) yeast tRNA (1 µg/µl final) 0.625 µl 1 M HEPES, pH 7.0 (25 mM final, pass through 0.22 µ filter prior to use to remove particulates) 0.75 µl 10 % SDS (0.3 % final) and DEPC-water to 25 µl final volume. The labelled cDNA target samples were filtered in Millipore 0.22 µ filter spin column (Ultrafree-15 MC, Millipore, USA) according to the manufacturer's instructions, followed by incubation of the reaction mixture at 100 °C for 2-5 minutes. The cDNA probes were cooled at room temp for 2-5 minutes by spinning at maximum speed in a microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was carefully placed on top of the SSA4 microarrays spotted on Immobilizer™ MicroArray Slide, and the hybridization mixture was applied to the array from 20 the side. An aliquot of 30 µL of 3xSSC was added to both ends of the hybridization chamber, and the slide was placed in the hybridization chamber (DieTech, USA). The chamber was sealed watertight and incubated at 65°C for 16-18 hours submerged in a water bath. After hybridization, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The slides were washed sequentially by plunging gently in 2 x 25 SSC/0.1% SDS at room temperature until the cover slip falls off into the washing solution, then in 1x SSC pH 7.0 (150 mM NaCl, 15 mM Sodium Citrate) at room temperature for 1 minute, then in 0.2 x SSC, pH 7.0 (30 mM NaCl, 3 mM Sodium Citrate) at room temperature for 1 minute, and finally in 0.05 x SSC (7.5 mM NaCl, 0.75 mM Sodium Citrate) for 5 seconds, followed by drying of the slides by spinning at 1000 xg for 2 minutes. The slides were stored in 30 a slide box in the dark until scanning.

### *Microarray Data Analysis*

The slides were scanned in a ScanArray 4000XL confocal laser scanner (Packard Instruments, USA). The hybridization data were analysed using the GenePix Pro 4.01 microarray analysis software (Axon, USA).

In the data analysis, the differences in labelling efficiency between the two fluorescent dyes were scaled by using an internal normalization approach. The average signal intensities from the control capture probes (Table 17) were used to calculate the normalization factor. This factor was multiplied to the signal intensity values from the Cy-3 target. Analysis of the 5 data demonstrates that capture probes with LNA in every third position have up to 5.2 fold higher signal-to-noise ratios, compared to the DNA capture probes (Fig. 39).

#### Example 11: Interpretation of Splice Array Data Using LNA Discriminating Probes.

This example illustrates the interpretation of microarray analysis of alternative mRNA 10 splicing. Different LNA capture probe design types are formalized, and the expression constant  $\emptyset$  is introduced as a measurement of alternative splicing.

##### *Introduction*

The eukaryotic pre-mRNA is the subject of Splicing and Alternative Splicing, hence sequences refer to RNA sequences, Original sequence refers to pre-mRNA, and splice forms 15 refer to mRNA sequences. The splicing is conducted by a cellular machinery named the spliceosome. The terms exons and introns can be used to refer to regions of pre-RNA sequences (or more specifically a single splice form). It is noted that a part of the corresponding DNA/pre-mRNA sequence that is an exon (not excised) in one splice form can potentially be absent in another splice form (e.g., partly absent in exon truncation and 20 completely absent in exon skipping). Thus, the terms "constant regions" and "variable regions" (see below) are useful for characterizing the process of identifying different splice forms.

Splicing can be defined as the production of a new sequence via the excision of part(s) of an original sequence (Fig. 40). Alternative splicing can be defined as the production of more than one novel sequence via the excision of different parts of the original sequence. When 25 comparing two different splice forms, they can be divided into a constant region that is shared by both sequences and a variable region by which the two splice forms differ (Fig. 41).

Alternative splicing can be categorized in terms of (i) whether or not the variable region is flanked by a single constant region or surrounded by two constant regions, (ii) the size of the variable region (e.g., exon skipping/intron retention vs. extension and truncation) [(intron/exon) 30 5' and 3'], and (iii) the number of variable regions (and hence the number of splice forms).

##### *Capture Probe Design*

Capture Probe design can be divided into 3 distinct types according to their position: Merged Probes (MP) or Junction Probes, Unique Internal Probes (UIP), and Shared Internal

Probes (SIP) (Fig. 42). Considering the case of a single variable region surrounded by constant regions, there are several different possible capture probe positions for each type (Fig. 43).

#### *Data Interpretation*

The aim of the analyses can be to determine (i) whether a given original sequence is subject to alternative splicing (i.e., whether there is more than one splice form present), and (ii) whether there is a difference in alternative splicing of the original sequence between two biological samples (i.e., whether the proportions between the two splice forms differ between biological samples). The analysis can also be used for data validation.

Possible biases in the microarray platform include (a) noise in terms of non-specific binding and subsequent false signal, (b) differences in dye labeling efficiency, (c) differences in capture probe affinity, (d) differences in sample conditions (e.g., number of cells, and amount of RNA), and (e) differences in reverse transcriptase efficiency of different splice forms. Biases can be corrected for by various means of normalization and/or standardization.

#### *Data Analysis*

In order to analyze the expression of the different splice forms, the expression constant  $\emptyset$  is introduced.  $\emptyset$  denotes the relation between the proportions of the signals (capture probes a and b) between the labeled extracts from biological samples (labeled with Cy5 & Cy3). That is,

$$(Cy5a / Cy3a) = (Cy5b / Cy3b) * \emptyset \quad \text{or,}$$

20

$$\begin{aligned} \emptyset &= (Cy5a / Cy3a) / (Cy5b / Cy3b) \\ &= (Cy5a * Cy3b) / (Cy5b * Cy3a) \text{ and,} \end{aligned}$$

$$\emptyset = (Cy5a / Cy5b) / (Cy3a / Cy3b)$$

25

$$\begin{aligned} &= (Cy5a * Cy3b) / (Cy3a * Cy5b) \\ &= (Cy5a * Cy3b) / (Cy5b * Cy3a) \text{ [same as above]} \end{aligned}$$

Considering normalization due to different biases and given a sample normalization factor S due to differences between the samples in terms of amounts of RNA, RT-efficiency, dye properties, etc. and a probe normalization factor P due to differences in probes in terms of affinity, position in target sequence, etc., the following equations apply.

For two probes: a and b,  $a * P = b$

For two samples Cy5 & Cy3,  $Cy5 * S = Cy3$ ,

35

Thus, considering two probes from two samples the signals are:

- Cy5a \* P \* S
- Cy5b \* S
- Cy3a \* P
- 5 - Cy3b

With respect to  $\emptyset$ :

$$\begin{aligned} \emptyset &= [(Cy5a * P * S) / (Cy3a * P)] / [(Cy5b * S) / (Cy3b)] \\ 10 \quad &= (Cy5a * P * S * Cy3b) / (Cy3a * P * Cy5b * S) \\ &= (Cy5a * Cy3b * P * S) / (Cy3a * Cy5b * P * S) \\ &= (Cy5a * Cy3b) / (Cy3a * Cy5b) * (P * S) / (P * S) \\ &= (Cy5a * Cy3b) / (Cy3a * Cy5b) * 1 \\ &= (Cy5a * Cy3b) / (Cy3a * Cy5b) \text{ [same as without normalization]} \end{aligned}$$

15

Note that the calculation of  $\emptyset$  is not affected by the normalization factors S and P, hence it is not necessary to normalize the array data when interpreting alternative splice arrays with the use of the Expression constant  $\emptyset$ .

#### *Properties of $\emptyset$*

- 20 If  $\emptyset = 1$ , there is no difference in the proportions of the targets of capture probes a and b in the two samples. Even in the case of alternative splicing, it is not possible to determine whether there is more than a single splice form present using this particular method. If  $\emptyset \neq 1$ , there is a difference in the proportions of the targets of capture probes a and b in the two samples, thus there is a difference in splice pattern and therefore there must be more than one
- 25 splice form present.

#### *Comparing $\emptyset$ 's*

$\emptyset$  can be compared between different transcripts to determine whether they have correlated expression, and  $\emptyset$ 's from sets of capture probes from the same transcript (different probes) can be averaged.

#### 30 *Example*

Considering a simple example of a single large variable region surrounded by constant regions using a combination of a Merged Probe and a Shared Internal Probe. Calculating  $\emptyset$  of a single splice form can be performed using the following equation:

$$35 \quad \emptyset = (Cy5MP * Cy3SIP) / (Cy3MP * Cy5SIP)$$



If  $\emptyset = 1$ , there is no difference in the proportions of the targets of capture probes a and b in the two samples, and it may not be possible to determine whether multiple splice forms are present using this particular method. If  $\emptyset \neq 1$ , there is a difference in the proportions of the targets of 5 capture probes a and b, thus there is a difference in splice pattern and therefore there must be more than one splice form present.

### *Conclusions*

It is possible to infer difference in expression level of two capture probe targets from two tissues when one is comparing the proportions of signals from one capture probe with the 10 proportion of signals from the other probe. In contrast, single signals may be subject to biases from normalizations and standardizations for each probe and sample.

### Example 12: Exemplary Microarrays

The nucleic acid arrays of the invention can be generated by standard methods for either 15 synthesis of nucleic acid probes that are then bonded to a solid support or synthesis of the nucleic acid probes on a solid support (e.g., by sequential addition of nucleotides to a reactive group on the solid support). In desirable methods for on-chip synthesis of the capture probes, photogenerated acids are produced in light-irradiate sites of the chip and used to deprotect the 5'-OH group of nucleic acid monomers and oligomers (e.g., to remove an acid-labile protecting 20 group such as 5'-O-DMT) to which a nucleotide is to be added (Gao et al., Nucleic Acid Research 29:4744-4750, 2001). Standard methods can also be used to label the nucleic acids in a test sample with, e.g., a fluorescent label, incubate the labeled nucleic acid sample with the array, and remove any unbound or weakly bound test nucleic acids from the array. Exemplary methods are described, for example, in U.S.P.N. 6,410,229; 6,406,844; 6,403,957; 6,403,320; 25 6,403,317; 6,346,413; 6,344,316; 6,329,143; 6,310,189; 6,309,831; 6,309,823; 6,261,776; 6,239,273; 6,238,862; 6,156,501; 5,945,334; 5,919,523; 5,889,165; 5,885,837; 5,744,305; 5,445,934; 5,800,9927; and 5,874,219.

In an exemplary method for synthesis of an array, capture probes were immobilized using AQ technology with a HEG5 linker (U.S.P.N. 6,033,784) onto an 30 Immobilizer™ slide. An exemplary chip consists of 288 spots in four replicates (i.e., 1152 spots) with a pitch of 250  $\mu\text{m}$ , and an exemplary hybridization buffer is 5xSSCT (i.e., 750 mM NaCl, 75 mM Sodium Citrate, pH 7.2, 0.05% Tween) and 10 mM  $\text{MgCl}_2$ . An exemplary target is a 45-mer oligonucleotide with Cy5 at the 5' end and with a final concentration in the hybridization solution of 1  $\mu\text{M}$ .

Hybridization was performed with 200  $\mu$ L hybridization solution in a hybridization chamber created by attaching a CoverWell™ gasket to the Immobilizer slide. The incubation was conducted overnight at 4°C. After hybridization, the hybridization solution was removed, and the chamber was flushed with 3 x 1.0 mL 5 hybridization buffer described above without any target nucleic acid. A coverWell™ chamber was then filled with 200  $\mu$ L hybridization solution without target. The slide was observed with a Zeiss Axioplan 2 epifluorescence microscope with a 5x Fluor objective and a Cy5 filterset from OMEGA. The temperature of the microscope stage was controlled with a Peltier element. Thirty-five images at each temperature were 10 acquired automatically with a Photometrics camera, automated shutter, and motorized microscope stage. The images were acquired, stitched together, calibrated and stored in stack by the software package "MetaVue"

Arrays can be generated using capture probes of any desired length (e.g., arrays of pentamers, hexamers, or heptamers.) In various embodiments, 1, 2, 3, 4, 5, 6, 7, 8, or more 15 nucleotides of the probes are LNA nucleotides. Desirably, at least 1, 2, 3, 5, 7, 9, or all of the A and T nucleotides in the probes are LNA A and LNA T nucleotides. LNA nucleotides can be placed in any position of the capture probe, such as at the 5' terminus, between the 5' and 3' termini, or at the 3' terminus. LNA nucleotides may be consecutive or may be separated by one or more other nucleotides. The microarrays can be used to analyze target nucleic acids of any 20 "AT" or "GC" content, and are especially useful for analyzing nucleic acids with high "AT" content because of the increased affinity of the microarrays of the present invention for such nucleic acids compared to traditional microarrays. Desirably, the array has at least 100, 200, 300, 400, 500, 600, 800, 1000, 2000, 5000, 8000, 10000, 15000, 20000, or more different probes. If desired, nucleotides with a universal base can be included in the capture probes to 25 increase the  $T_m$  of the capture probes (e.g., capture probes of less than 7, 6, 5, or 4 nucleotides). Exemplary "non-discriminatory" nucleotides include inosine, random nucleotides, 5 nitro-indole, LNA, inosine, and LNA 2-aminopurine. In desirable embodiments, 1, 2, 3, 4, 5, or more nucleotides with a universal base are located at the 5' and/or 3' termini of the capture probes.

30

#### Example 13: Exemplary Application of Nucleic Acids of the Invention

An exemplary application of these methods includes comparing hybridization patterns of cDNA or cRNA from a patient sample to classify early-tumors or detect an infection or a diseased state. The microarrays of the invention may also be used as a 35 general tool to analyze the PCR products generated by amplification of a test sample

with PCR primers for one or more nucleic acids of interest. For example, PCR primers can be used to amplify nucleic acids with a particular exon or exon-exon combination, and then the PCR products can be identified and/or quantified using a microarray of the invention. For identification of splice variants, PCR primers to specific exons can be used to amplify nucleic acids that are then applied to a microarray for detection and/or quantification as described herein. To detect microbial pathogens, species-specific PCR primers (e.g., primers specific for an exon whose sequence differs among species) can be used to amplify nucleic acids in a sample for subsequent analysis using a microarray. For example, the hybridization pattern of the PCR products to the array can be used to distinguish between different bacteria, viruses, or yeast and even between different strains of the same pathogenic species. In particular embodiments, the array is used to determine whether a patient sample contains a bacteria strain that is known to be resistant or susceptible to particular antibiotics or contains a virus or yeast strain known to be resistant or susceptible to certain drugs. Changes in product composition or raw material origin can also be detected using a microarray. The arrays can also be used to determine the composition of mRNA cocktails.

Exemplary environmental microbiology applications of these arrays include identification of major rRNA types in contaminated soil samples and classification of microbial isolates. These rRNA amplicates are formed from rRNA by rtPCR or from the rDNA gene by conventional PCR. Numerous general and selective primers for different groups of organisms have been published. Most frequently an almost full length amplicate of the 16S rDNA gene is used (e.g., the primers 26F and 1492R). For purifying rRNA from a soil sample, standard methods such as one or more commercial extraction kits from companies such as QIAGEN ("Rneasy", Q-biogene "RNA PLUS," 25 or "Total RNA safe" can be used.

#### Example 14: Methods for Minimizing the Variance in Melting Temperatures in Nucleic Acid Populations of the Invention

Any simultaneous use of more than one primer or probe is made difficult because the involved primers or probes must work under the same conditions. An indication of whether or not two or more primers or probes will work under the same conditions is the relative  $T_{ms}$  at which the hybridized oligonucleotides dissociate. In cases where probes are applied for specific detection of homologous sequences such as splice variants, the  $\Delta T_m$  is of importance.  $\Delta T_m$  expresses the difference between  $T_m$  of the match and the  $T_m$  of the mismatch hybridizations. Generally, the larger  $\Delta T_m$  obtained, the more specific detection of the sequence of interest. In

addition, a large  $\Delta T_m$  facilitates more probes to be used simultaneously and in this way a higher degree of multiplexity can be applied.

High affinity nucleotide analogs such as LNA can be also be used universally to equalize the melting properties of oligonucleotides with different AT and CG content. The increased affinity of LNA adenosine and LNA thymidine corresponds approximately to the normal affinity of DNA guanine and DNA cytosine. An overall substitution of all DNA-A and DNA-T with LNA-A and LNA-T results in melting properties that are nearly sequence independent but only depend on the length of the oligonucleotide. This may be important for design of oligonucleotide probes used in large multiplex analysis. The effect of LNA A and T substitutions has been evaluated by predicting the  $T_m$  value of all possible 9-mer oligonucleotides with different universal substitutions. The distribution of the 262,000  $T_m$ -values exhibits a very homogeneous  $T_m$  value for universally LNA A and T substituted oligonucleotides. The standard deviation of the melting temperature for all 9-mers drops from 7.7°C for pure DNA to only 2.2°C for LNA A and T substituted oligonucleotides. This equalizing effect may also be utilized for photomediated on-chip synthesis of oligonucleotides.

It is often difficult to design probes and primers with the same range of melting temperature due to the variance in A/T and G/C content of the probing sites. Highly A/T rich regions typically give lower  $T_m$  values. Furthermore, if single mismatches are to be resolved, G/T mismatches are known to contribute little to  $\Delta T_m$ . As discussed above, the use of LNA is a desirable way to solve problems related to multiplex use of primers and probes. LNA offers the possibility to adjust  $T_m$  and increase the  $\Delta T_m$  at the same time. LNA increases  $T_m$  with 4-8°C/substitution and increases  $\Delta T_m$  in many cases (Table 9).

Table 9. Demonstration of LNA controlled increase of  $T_m$  and  $\Delta T_m$ .

$T_m$ of LNA:DNA Duplexes	Perfect match 3'-ACGACCAC-5'	Single mismatch 3'-ACGGCCAC-5'	$\Delta T_m$
LNA 8-mer 5'-TGCIGGTG-3'	71°C	45°C	26°C
DNA 8-mer 5'-TGCIGGTG-3'	35°C	25°C	10°C

As LNA can be mixed with DNA during standard oligonucleotide synthesis, LNA can be placed at optimal positions in probes in order to adjust  $T_m$ . The specificity of PCR may also be enhanced by the use of LNA in primers, or probes, and this facilitates a higher degree of multiplexity. By incorporation of LNA, the  $T_m$  of the primers or probes can be adjusted to

work at the same temperature. Amplification or hybridization is more specific when LNA is included in primers or probes. This is due to the LNA increased  $\Delta T_m$ , which relates to higher specificity. Once  $\Delta T_m$  of the primers or probes is high, more primers or probes can potentially be brought to work together.

### 5 Prediction of $T_m$

LNA can be used to enhance any experiment that is based on hybridization. The series of algorithms described herein have been developed to predict the optimal use of LNA. Melting properties of 129 different LNA substituted capture probes hybridized described herein to their corresponding DNA targets were measured in solution using UV-spectrophotometry.

10 The data set was divided into a training set with 90 oligonucleotides and a test set with 39 oligonucleotides. The training set was used for training of both linear regression models and neural networks. Neural networks trained with nearest neighbour information, length, and DNA/LNA neighbour effect are efficient for prediction of  $T_m$  with the given set of data.

*Applications of the Normalization of Thermal Stability by LNA A and T Nucleotide*

### 15 Substitutions

All assays in which DNA/RNA hybridization is conducted may benefit from the use of LNA in terms of increased specificity and quality. Exemplary uses include sequencing, primer extension assays, PCR amplification, such as multiplex PCR, allele specific PR amplification, molecular beacons, (e.g., nucleic acids be multiplexed with one colour based on multiple  $T_m$ 's),

20 Taq-man probes, *in situ* hybridization probes (e.g., chromosomal and bacterial 16S rRNA probes), capture probes to the mRNA poly-A tail, capture probes for microarray detection of SNPs, capture probes for expression microarrays (sensitivity increased 5-8 times), and capture probes for assessment of alternative mRNA splicing.

### 25 Example 15: Exemplary Methods for the Prediction of Melting Temperatures for Nucleic Acid Populations of the Invention

LNA units have different melting properties than DNA and RNA nucleotides. Until recently, thermodynamical models for melting temperature prediction have existed for DNA and RNA only, but not for LNA. Now a  $T_m$  prediction model for LNA/DNA mixed  
30 oligonucleotides has been developed. The  $T_m$  prediction tool is available on-line at the Exiqon website ([www.LNA-Tm.com](http://www.LNA-Tm.com) and <http://www.exiqon.com/Poster/Tmpred-ET-view.pdf>).

Numerous applications in molecular biology are based on the ability of DNA and RNA to hybridize in a temperature dependent manner (e.g. the microarray techniques, PCR reactions and blotting techniques). The melting properties of nucleic acid duplexes, in particular the  
35 melting temperature  $T_m$ , are crucial for optimal design of such experiments.  $T_m$  is usually

computed using a two-state thermodynamical model (Breslauer, *Meth. Enzymol.*, 259:221-242, 1995). Several different groups have estimated model parameters for nearest neighbors in the sequence based on experimental data (for a review see SantaLucia, *Proc. Natl. Acad. Sci.*, 95:1460-1465, 1998).

5 The model described herein predicts the  $T_m$  of duplexes of mixed LNA/DNA oligonucleotides hybridized to their complementary DNA strands. DNA monomers are denoted with lowercase letters, and LNA monomers are denoted with uppercase letters, e.g., there are eight types of monomers in the mixed strand: a, c, g, t, A, C, G and T. The model is based on the formula (SantaLucia, 1998, *supra*; Allawi *et al.*, *Biochemistry* 36:10581-10594, 1997).

10

$$T_m = \frac{\Delta H}{\Delta S + R \cdot \ln(C - C_m/2) + 0.368(L-1)\ln[Na^+]}$$

in which the salt concentration  $[Na^+]$  enters as an entropic correction together with the oligonucleotide concentrations.  $R$  is the gas constant,  $C$  and  $C_m$  are the concentrations of the  
15 two strands where  $C \geq C_m$ , and  $L$  is the length of the strands. For self-complementary sequences,  $C - C_m/2$  is replaced by the total strand concentration  $C_T$  and a symmetry correction of  $-1.4$  cal/k·mol is added to  $\Delta S$  (SantaLucia, 1998, *supra*).

The LNA model differs from SantaLucia's DNA model in the way the changes in enthalpy  $\Delta H$  and entropy  $\Delta S$  are calculated. As in SantaLucia's model, they depend on nearest  
20 neighbor sequence information and special contributions for the terminal base-pairs in the two ends of the duplex. However, with eight types of monomers (LNA and DNA) the increased number of nearest neighbor combinations requires more model parameters to be determined and hence more data.

#### *Parameter Reduction*

25 Usually  $\Delta H$  and  $\Delta S$  are calculated as a sum of contributions from all nearest neighbor pairs in the sequence. The inclusion of LNA doubles the number of monomer types and quadruples the number of possible nearest neighbor pairs. Parameter reduction strategies are used for matching the model complexity to limited data sets. A strategy for reducing model complexity is to sum  $\Delta H$  from single base-pair contributions, which do not take the influence of  
30 adjacent nucleotides into account. However, nearest neighbor contributions are added as a correction term to the single base-pair contributions.

Another strategy is to use hierarchically reduced monomer alphabets. Here, similar monomers are identified with the same letter. A four-letter alphabet, {w,s,W,S}, defines classes

according to binding strength:  $w=\{a,t\}$ ,  $s=\{c,g\}$ ,  $W=\{A,T\}$  and  $S=\{C,G\}$ . The smallest alphabet,  $\{D,L\}$ , simply identifies the monomer type: DNA or LNA. As an example, the sequence GcTAAcTt can be written as SsWWWsWw or as LDLLLDLD.

The principle is to split  $\Delta H$  and  $\Delta S$  into contributions that depend on different levels of detail of the sequence. The fine levels of detail require many parameters to be determined, while the coarse levels need fewer parameters. The more detailed contributions can then be treated as minor corrections, thus effectively reducing the total number of model parameters.

### Training

Model parameters were determined using data from melting experiments on hundreds of oligonucleotides. The oligonucleotides were random sequences with lengths between 8 and 20 and a percentage of LNA between 20 and 70. Melting curves were obtained using a Perkin-Elmer UV  $\lambda$ -40 spectrophotometer, but only the  $T_m$  values were used for modeling. Model parameters were adjusted using a gradient descent algorithm that minimizes the error function

$$E = \sum_{\substack{\text{data} \\ \text{set}}} \frac{1}{N} (T_m^{\text{pred}} - T_m^{\text{exp}})^2,$$

i.e., the distance between predicted and experimental  $T_m$  values. Many different models were trained in this way and their performance was evaluated on test sets distinct from the training data. Seven reliable models were chosen and combined to form the committee model implemented at the Exiqon website ([www.LNA-Tm.com](http://www.LNA-Tm.com).)

### Machine Learning And Thermodynamics

The aim of this work has been to estimate  $T_m$  values as accurately as possible. To this end, a machine learning approach has been adopted in which the prediction of the physical  $\Delta H$  and  $\Delta S$  quantities is less important. The parameters of this model may be inaccurate as thermodynamic quantities. First, the gradient descent algorithm produces a broad ensemble of models in which the  $\Delta H$  and  $\Delta S$  parameters can vary substantially, while maintaining an accuracy in the predicted  $T_m$ . Second, the thermodynamic meaning of  $\Delta H$  and  $\Delta S$  is based on a two-state assumption, which may not be realistic in every case. Even short oligonucleotides can form different secondary structures or melt through multiple-state transitions (Tøstesen *et al.*, J. Phys. Chem. B. 105:1618-1630, 2001). Third, the use of an optical instrument instead of a calorimetric instrument (DSC) introduces an error in the measured  $\Delta H$  and  $\Delta S$ . Nevertheless, the uncertain thermodynamic interpretation of the  $\Delta H$  and  $\Delta S$  model parameters does not imply that the  $T_m$  prediction model is unreliable.

### Results

The  $T_m$  prediction model has been tested on two data sets that were not used during the training process. One set consisted of pure DNA oligonucleotides without LNA monomers and had a standard deviation of the residuals (SEP) of 1.57 degrees. The other set consisted of mixed oligonucleotides with both LNA and DNA and had a SEP of 5.25 degrees. The difference in prediction accuracy between the two types of oligonucleotides suggests that  $T_m$  prediction of mixed strands is a more complex task than  $T_m$  prediction of pure DNA. This is possibly due to irregularities in the duplex helical structure induced by the LNA monomers (Nielsen *et al.*, Bioconjug. Chem. 11:228-238, 2000). The obtained prediction accuracy is in both cases adequate for most biological applications. In conclusion, the reduced nearest neighbor model implemented at the Exiqon website ([www.LNA-Tm.com](http://www.LNA-Tm.com)) can predict  $T_m$  surprisingly well for both types of oligonucleotides. This indicates that the parameter reduction strategy is applicable for other types of modified oligonucleotides.

Example 16: Optional Algorithm to Optimize the Substitution Pattern of Nucleic Acids of the  
15 Invention

High affinity nucleotides such as LNA and other nucleotides that are conformationally restricted to prefer the C3'-endo conformation or nucleotides with a modified backbone and/or nucleobase stabilize a double helix configuration. As these effects are generally additive, the most stable duplex between a high affinity capture oligonucleotide and an unmodified target oligonucleotide should generally arise when all nucleotides in the capture probe or primer are replaced by their high affinity analogue. The most stable duplex should thus be formed between a fully modified LNA capture probe and the corresponding DNA/RNA target molecule. Such a fully modified capture probe should be more efficient in capturing target molecules, and the resulting duplex is more thermally stable.

25 However, many high affinity nucleotides (e.g., as LNA) have an even higher affinity for other high affinity nucleotides (e.g., as LNA) than for DNA/RNA. A fully modified capture probe may thus form duplexes with itself, or if it is long enough, internal hairpins that are even more stable than duplexes with the desired target molecule. Probes with even a small inverse repeat segment where all constituent positions are substituted with high affinity nucleotides 30 may bind to itself and be unable to bind the target. Thus, a sequence dependent substitution pattern is desirably used to avoid substitutions in positions that may form self-complementary base-pairs.

For example, a computer algorithm can be used to automatically determine the optimal substitution pattern for any given capture probe sequence according to the following two 35 criteria. First, the difference between the stability of (i) the duplex formed between the capture



probe and the target molecule and (ii) the best possible duplex between two capture probes should be above a certain threshold. If this is not possible, then the substitution pattern with the largest possible difference is chosen. Second, the capture probe should contain as many substitutions as possible in order to bind as much target as possible at any given temperature and to increase the thermal stability of the formed duplex. Alternatively, the second criterion is substituted with the following alternative criterion to obtain capture probes with similar thermal stability. The number and position of capture probe substitutions should be adjusted so that all the duplexes between capture probes and targets have a similar thermal stability (i.e.,  $T_m$  equalization).

10 For oligonucleotide capture probes such, incomplete matches between target and capture probe are likely to be a reproducible feature of the recorded biosignatures. For short probes, the second criterion for increasing thermal stability is more desirable than the alternative second criterion for  $T_m$  equalization. For long capture probes and PCR primers, the second alternative criterion is desirably used since  $T_m$  equalization is desirable for these probes and  
15 primers.

An exemplary algorithm works as follows. For each nucleotide sequence in an array of length  $n$ , all possible substitution patterns, i.e.,  $2^n$  different sequences are evaluated. Each evaluation consists of estimating the energetic stability of the duplex between the substituted capture sequence and a perfect match unmodified target ("target duplex") and the energetic  
20 stability of the most stable duplex that can be formed between two substituted capture probes themselves ("self duplex").

The energetic stability estimate for a duplex may be calculated, e.g., using a Smith-Waterman algorithm with the following scoring matrix.

Gap initiation penalty: -8  
25 Gap continuation penalty: -50

	a	c	g	t	A	C	G	T
a	-2							
c	-2	-2						
g	-2	3	-2					
30 t	2	-2	1	-2				
A	-3	-3	-3	4	-3			
C	-3	-3	6	-3	-3	-3		
G	-3	6	-3	2	-3	9	-3	
35 T	4	-3	2	-3	6	-3	3	-3

This scoring matrix was partly based on the best parameter fit to a large (over 1000) number of melting curves of different DNA and LNA containing duplexes and partly by visual scoring of test capture probe efficiency. If desired, this scoring matrix may be optimized by optimizing the parameter fit as well as increasing or optimizing the dataset used to obtain these parameters.

- 5 As an example of these calculations, the heptamer sequence ATGCAGA in which each position can be either an LNA or a DNA nucleotide is used. The target duplex formed between a fully modified capture probes with this sequence and its unmodified target receive a score of 34 as illustrated below.

10      Capture sequence:    A-T-G-C-A-G-A  
                                   | | | | | | |  
             Target sequence:    t-a-c-g-t-c-t  
             Score:                    4+4+6+6+4+6+4 = 34

- 15 The most stable self duplex that can be formed between two modified capture probes has an almost equivalent energetic stability with a score of 30 as illustrated below.

            Capture sequence:    A-T-G-C-A-G-A  
                                   | | | |  
 20      Target sequence:    A-G-A-C-G-T-A  
             Score:                    +6+9+9+6            = 30

- Thus, the capture probe efficiency of a fully modified probe is likely reduced by its propensity to form a stable duplex with itself. In contrast, by choosing a slightly different substitution  
 25 pattern, ATGcaGA in which capital letters represent LNA nucleotides, the stability of the target duplex is reduced slightly from 34 to 29.

            Capture sequence:    A-T-G-c-a-G-A  
                                   | | | | | | |  
 30      Target sequence:    t-a-c-g-t-c-t  
             Score:                    4+4+6+3+2+6+4 = 29

However, the most stable self complementary duplex that can be formed is reduced much more from 30 to 20, as illustrated below.

35

122

Capture sequence:    A-T-G-c-a-G-A

| | | |

Target sequence:    A-G-a-c-G-T-A

Score:                    +4+6+6+4       = 20

5

The difference between the stability of the desired target duplex and the undesired self duplex can be further increased by using the capture sequence AtgcaGA where the target duplex has a score of 24.

10      Capture sequence:    A-t-g-c-a-G-A

| | | | | | |

Target sequence:    t-a-c-g-t-c-t

Score:                    4+2+3+3+2+6+4 = 24

15 Whereas the score of the self duplex is only 10, as shown below.

Capture sequence:    A-t-g-c-a-G-A

| | | |

Target sequence:    A-G-a-c-g-t-A

20      Score:                    +2+3+3+2       = 10

The additional destabilization of the self duplex is generally not required if the difference in stability between the target duplex and self duplex is above a threshold of 25% of the target duplex stability, as illustrated below.

25

Discrimination for ATGCAGA =  $(34-30)/34 = 12\% < \text{threshold}$   
(25%)

Discrimination for ATGcaGA =  $(29-20)/29 = 31\% \geq \text{threshold}$   
(25%)

30

Discrimination for ATGCAGA =  $(24-10)/24 = 58\% \geq \text{threshold}$   
(25%)

Thus, ATCCaGA is the substitution pattern with the highest degree of substitution for which the stability of the target duplex is adequately more stable than the stability of the best self duplex 35 (e.g., above 25%).

This algorithm can be used to determine desirable substitution patterns for any size capture probe or any given probe sequence. The following simple design rules may also be applied for probe design, especially for short probes. The best self alignment for the corresponding DNA capture probe in the sequence is determined using a simple Smith-5 Waterman scoring matrix of:

	a	c	g	t
a	-2			
c	-2	-2		
10 g	-2	3	-2	
t	2	-2	1	-2

Additionally, all possible positions in the sequence are substituted, with the exception of desirably avoiding the substitution of both bases of a self-complementary base-pair. The most 15 stable self duplex thus does not contain any LNA:LNA base-pairs but only LNA:DNA basepairs.

#### *Exemplary Computer*

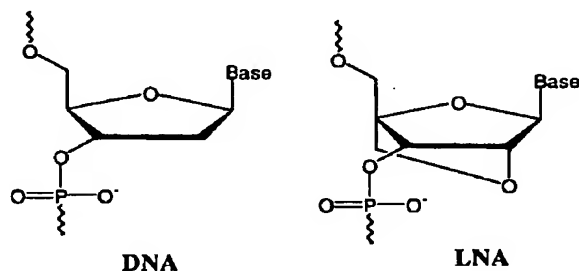
Any of the methods described herein may be implemented using virtually any computer.

Fig. 44 shows such an exemplary computer system. Computer system 2 includes internal and 20 external components. The internal components include a processor 4 coupled to a memory 6. The external components include a mass-storage device 8, *e.g.*, a hard disk drive, user input devices 10, *e.g.*, a keyboard and a mouse, a display 12, *e.g.*, a monitor, and usually, a network link 14 capable of connecting the computer system to other computers to allow sharing of data and processing tasks. Programs are loaded into the memory 6 of this system 2 during operation. 25 These programs include an operating system 16, *e.g.*, Microsoft Windows, which manages the computer system, software 18 that encodes common languages and functions to assist programs that implement the methods of this invention, and software 20 that encodes the methods of the invention in a procedural language or symbolic package. Languages that can be used to program the methods include, without limitation, Visual C/C++ from Microsoft. In preferred 30 applications, the methods of the invention are programmed in mathematical software packages that allow symbolic entry of equations and high-level specification of processing, including algorithms used in the execution of the programs, thereby freeing a user of the need to program procedurally individual equations or algorithms. An exemplary mathematical software package useful for this purpose is Matlab from Mathworks (Natick, MA). Using the Matlab software, 35 one can also apply the Parallel Virtual Machine (PVM) module and Message Passing Interface

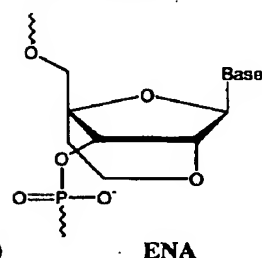
(MPI), which supports processing on multiple processors. This implementation of PVM and MPI with the methods herein is accomplished using methods known in the art. Alternatively, the software or a portion thereof is encoded in dedicated circuitry by methods known in the art.

### 5 Example 17.: Exemplary Locked Nucleic Acids (LNA)

As disclosed in WO 99/14226, LNA are DNA analogues that form DNA- or RNA-heteroduplexes with exceptionally high thermal stability. LNA units include bicyclic compounds as shown immediately below where ENA refers to 2'O,4'C-ethylene-bridged



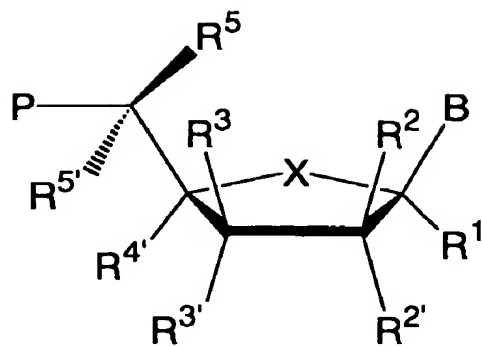
nucleic acids:



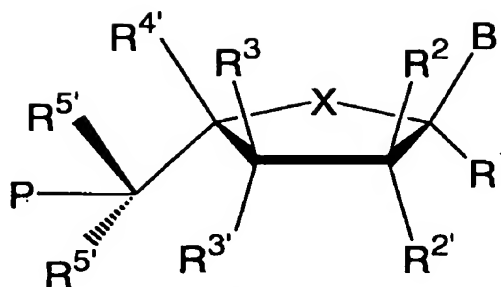
References herein to Locked Nucleoside Analogues, LNA units, LNA monomers, or similar terms are inclusive of such compounds as disclosed in WO 99/14226, WO 00/56746, WO 00/56748, and WO 00/66604.

- 15            Desirable LNA monomers and oligomers share some chemical properties of DNA and RNA; they are water soluble, can be separated by agarose gel electrophoresis, and can be ethanol precipitated.

Desirable LNA monomers and oligonucleotide units include nucleoside units having a 2'-4' cyclic linkage, as described in the International Patent Application WO 99/14226 and WO 20 0056746, WO 0056748, and WO 0066604. Desirable LNA monomers structures are exemplified in the formulae Ia and Ib below. In formula Ia the configuration of the furanose is denoted D -  $\beta$ , and in formula Ib the configuration is denoted L -  $\alpha$ . Configurations which are composed of mixtures of the two, e.g. D -  $\beta$  and L -  $\alpha$ , are also included.



Ia



Ib

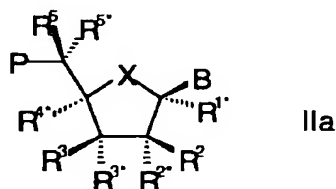
In Ia and Ib, X is oxygen, sulfur and carbon; B is a universal or modified base (particularly non-natural occurring base) e.g. pyrene and pyridyloxazole derivatives, pyrenyl, pyrenylmethylglycerol moieties, all of which may be optionally substituted. Other desirable universal bases include, pyrrole, diazole or triazole moieties, all of which may be optionally substituted, and other groups e.g. modified adenine, cytosine, 5-methylcytosine, isocytosine, pseudoisocytosine, guanine, thymine, uracil, 5-bromouracil, 5-propynyluracil, 5-propynyl-6-fluoroluracil, 5-methylthiazoleuracil, 6-aminopurine, 2-aminopurine, inosine, diaminopurine, 7-propyne-7-deazaadenine, 7-propyne-7-deazaguanine. R<sup>1</sup>, R<sup>2</sup> or R<sup>2'</sup>, R<sup>3</sup> or R<sup>3'</sup>, R<sup>5</sup> and R<sup>5'</sup> are 10 hydrogen, methyl, ethyl, propyl, propynyl, aminoalkyl, methoxy, propoxy, methoxy-ethoxy, fluoro, or chloro.

P designates the radical position for an internucleoside linkage to a succeeding monomer, or a 5'-terminal group, R<sup>3</sup> or R<sup>3'</sup> is an internucleoside linkage to a preceding monomer, or a 3'-terminal group. The internucleotide linkage may be a phosphate, 15 phosphorothioate, phosphorodithioate, phosphoramidate, phosphoroselenoate, phosphorodiselenoate, alkylphosphotriester, or methyl phosphonate. The internucleotide linkage may also contain non-phosphorous linkers, hydroxylamine derivatives (e.g. -CH<sub>2</sub>-NCH<sub>3</sub>-O-CH<sub>2</sub>-), hydrazine derivatives, e.g. -CH<sub>2</sub>-NCH<sub>3</sub>-NCH<sub>3</sub>-CH<sub>2</sub>, amid derivatives, e.g. -CH<sub>2</sub>-CO-NH-CH<sub>2</sub>-, CH<sub>2</sub>-NH-CO-CH<sub>2</sub>-. In Ia, R<sup>4'</sup> and R<sup>2'</sup> together designate -CH<sub>2</sub>-O-, -CH<sub>2</sub>-S-, 20 -CH<sub>2</sub>-NH-, -CH<sub>2</sub>-NMe-, -CH<sub>2</sub>-CH<sub>2</sub>-O-, -CH<sub>2</sub>-CH<sub>2</sub>-S-, -CH<sub>2</sub>-CH<sub>2</sub>-NH-, or -CH<sub>2</sub>-CH<sub>2</sub>-NMe- where the oxygen, sulfur or nitrogen, respectively, is attached to the 2'-position. In Formula Ib, R<sup>4'</sup> and R<sup>2</sup> together designate -CH<sub>2</sub>-O-, -CH<sub>2</sub>-S-, -CH<sub>2</sub>-NH-, -CH<sub>2</sub>-NMe-, -CH<sub>2</sub>-CH<sub>2</sub>-O-, -CH<sub>2</sub>-CH<sub>2</sub>-S-, -CH<sub>2</sub>-CH<sub>2</sub>-NH-, or -CH<sub>2</sub>-CH<sub>2</sub>-NMe- where the oxygen, sulphur or nitrogen, respectively, is attached to the 2-position (R<sup>2</sup> configuration).

25 Desirable LNA monomer structures are structures in which X is oxygen (Formula Ia and Ib); B is a universal base such as pyrene; R<sup>1</sup>, R<sup>2</sup> or R<sup>2'</sup>, R<sup>3</sup> or R<sup>3'</sup>, R<sup>5</sup> and R<sup>5'</sup> are hydrogen; P is a phosphate, phosphorothioate, phosphorodithioate, phosphoramidate, and methyl phosphonates;

$R^3$  or  $R^{3'}$  is an internucleoside linkage to a preceding monomer, or a 3'-terminal group. In Formula Ia,  $R^{4'}$  and  $R^{2'}$  together designate  $-\text{CH}_2-\text{O}-$ ,  $-\text{CH}_2-\text{S}-$ ,  $-\text{CH}_2-\text{NH}-$ ,  $-\text{CH}_2-\text{NMe}-$ ,  $-\text{CH}_2-\text{CH}_2-\text{O}-$ ,  $-\text{CH}_2-\text{CH}_2-\text{S}-$ ,  $-\text{CH}_2-\text{CH}_2-\text{NH}-$ , or  $-\text{CH}_2-\text{CH}_2-\text{NMe}-$  where the oxygen, sulphur or nitrogen, respectively, is attached to the 2'-position, and in Formula Ib,  $R^{4'}$  and  $R^{2'}$  together designate  $-\text{CH}_2-\text{O}-$ ,  $-\text{CH}_2-\text{S}-$ ,  $-\text{CH}_2-\text{NH}-$ ,  $-\text{CH}_2-\text{NMe}-$ ,  $-\text{CH}_2-\text{CH}_2-\text{O}-$ ,  $-\text{CH}_2-\text{CH}_2-\text{S}-$ ,  $-\text{CH}_2-\text{CH}_2-\text{NH}-$ , or  $-\text{CH}_2-\text{CH}_2-\text{NMe}-$  where the oxygen, sulphur or nitrogen, respectively, is attached to the 2'-position in the  $R^2$  configuration.

Particularly desirable LNA monomer for incorporation into an oligonucleotide of the invention include those of the following formula IIa



wherein X oxygen, sulfur, nitrogen, substituted nitrogen, carbon and substituted carbon, and desirably is oxygen; B is a modified base as discussed above e.g. an optionally substituted carbocyclic aryl such as optionally substituted pyrene or optionally substituted pyrenylmethylglycerol, or an optionally substituted heteroalicyclic or optionally substituted heteroaromatic such as optionally substituted pyridyloxazole. Other desirable universal bases include, pyrrole, diazole or triazole moieties, all of which may be optionally substituted;  $R^{1*}$ ,  $R^2$ ,  $R^3$ ,  $R^5$  and  $R^{5*}$  are hydrogen; P designates the radical position for an internucleoside linkage to a succeeding monomer, or a 5'-terminal group,  $R^{3*}$  is an internucleoside linkage to a preceding monomer, or a 3'-terminal group; and  $R^{2*}$  and  $R^{4*}$  together designate  $-\text{O}-\text{CH}_2-$  or  $-\text{CH}_2-\text{CH}_2-\text{O}-$  where the oxygen is attached in the 2'-position, or a linkage of  $-(\text{CH}_2)_n-$  where n is 2, 3 or 4, desirably 2, or a linkage of  $-\text{S}-\text{CH}_2-$  or  $-\text{NH}-\text{CH}_2-$ .

LNA units of formula IIa where  $R^{2*}$  and  $R^{4*}$  contain oxygen are sometimes referred to herein as "oxy-LNA"; units of formula IIa where  $R^{2*}$  and  $R^{4*}$  contain sulfur are sometimes referred to herein as "thio-LNA"; and units of formula IIa where  $R^{2*}$  and  $R^{4*}$  contain nitrogen are sometimes referred to herein as "amino-LNA". For many applications, oxy-LNA units are desirable modified nucleic acid units of oligonucleotides of the invention.

Particularly desirable LNA monomers for use in oligonucleotides of the invention are 2'-deoxyribonucleotides, ribonucleotides, and analogues thereof that are modified at the 2'-

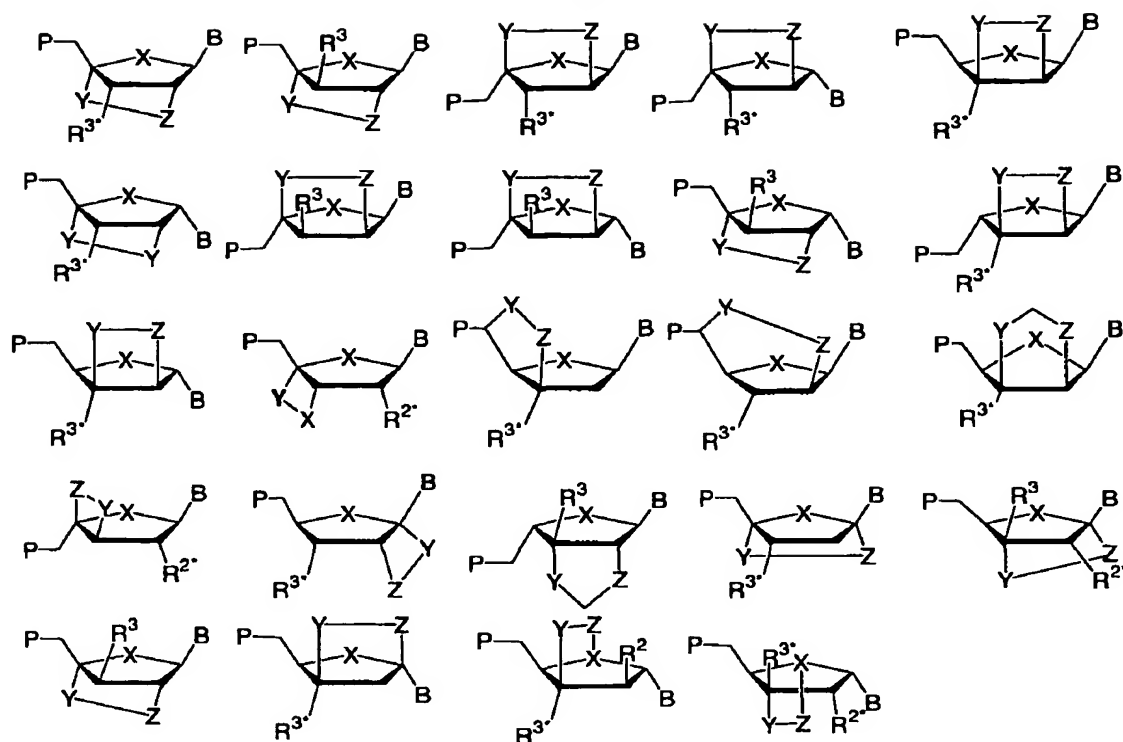
position in the ribose, such as 2'-O-methyl, 2'-fluoro, 2'-trifluoromethyl, 2'-O-(2-methoxyethyl), 2'-O-aminopropyl, 2'-O-dimethylamino-oxyethyl, 2'-O-fluoroethyl or 2'-O-propenyl, and analogues wherein the modification involves both the 2' and 3' position, desirably such analogues wherein the modifications links the 2'- and 3'-position in the ribose, such as those described in Nielsen et al., J. Chem. Soc., Perkin Trans. 1, 1997, 3423-33, and in WO 99/14226, and analogues wherein the modification involves both the 2'- and 4'-position, desirably such analogues wherein the modifications links the 2'- and 4'-position in the ribose, such as analogues having a -CH<sub>2</sub>-S- or a -CH<sub>2</sub>-NH- or a -CH<sub>2</sub>-NMe- bridge (see Singh et al. J. Org. Chem. 1998, 6, 6078-9). Although LNA monomers having the  $\beta$ -D-ribo configuration are often the most applicable, other configurations also are suitable for purposes of the invention. Of particular use are  $\alpha$ -L-ribo, the  $\beta$ -D-xylo and the  $\alpha$ -L-xylo configurations (see Beier et al., Science, 1999, 283, 699 and Eschenmoser, Science, 1999, 284, 2118), in particular those having a 2'-4' -CH<sub>2</sub>-S-, -CH<sub>2</sub>-NH-, -CH<sub>2</sub>-O- or -CH<sub>2</sub>-NMe- bridge.

In another desirable embodiment, LNA modified oligonucleotides used in this invention comprises oligonucleotides containing at least one LNA monomeric unit of the general scheme A above, wherein X, B, P are defined as above. One of the substituents R<sup>2</sup>, R<sup>2\*</sup>, R<sup>3</sup>, and R<sup>3\*</sup> is a group P\* which designates an internucleoside linkage to a preceding monomer, or a 2'/3'-terminal group. Two of the substituents of R<sup>1\*</sup>, R<sup>2</sup>, R<sup>2\*</sup>, R<sup>3</sup>, R<sup>4\*</sup>, R<sup>5</sup>, R<sup>5\*</sup>, R<sup>6</sup>, R<sup>6\*</sup>, R<sup>7</sup>, and R<sup>7\*</sup> when taken together designate a biradical structure selected from -(CR<sup>\*</sup>R<sup>\*</sup>)<sub>r</sub>-M-(CR<sup>\*</sup>R<sup>\*</sup>)<sub>s</sub>-, -(CR<sup>\*</sup>R<sup>\*</sup>)<sub>r</sub>-M-(CR<sup>\*</sup>R<sup>\*</sup>)<sub>s</sub>-M-, -M-(CR<sup>\*</sup>R<sup>\*</sup>)<sub>r+s</sub>-M-, -M-(CR<sup>\*</sup>R<sup>\*</sup>)<sub>r</sub>-M-(CR<sup>\*</sup>R<sup>\*</sup>)<sub>s</sub>-, -(CR<sup>\*</sup>R<sup>\*</sup>)<sub>r+s</sub>-, -M-, -M-M-, wherein each M is independently selected from -O-, -S-, -Si(R<sup>\*</sup>)<sub>2</sub>-, -N(R<sup>\*</sup>)-, >C=O-, C(=O)-N(R<sup>\*</sup>)-, and -N(R<sup>\*</sup>)-C(=O)-. Each R<sup>\*</sup> and R<sup>1(1\*)</sup>-R<sup>7(7\*)</sup>, which are not involved in the biradical, are independently selected from hydrogen, halogen, azido, cyano, nitro, hydroxy, mercapto, amino, mono- or di(C<sub>1-6</sub>-alkyl)amino, optionally substituted C<sub>1-6</sub>-alkoxy, optionally substituted C<sub>1-6</sub>-alkyl, DNA intercalators, photochemically active groups, thermochemically active groups, chelating groups, reporter groups, and ligands, and/or two adjacent (non-geminal) R<sup>\*</sup> may together designate a double bond, and each of r and s is 0-4 with the proviso that the sum r+s is 1-5.

Examples of LNA units are shown scheme B:

30





Scheme B

wherein the groups, X and B are defined as above. P designates the radical position for an internucleoside linkage to a succeeding monomer, nucleoside such as an L-nucleoside, or a 5'-terminal group, such internucleoside linkage or 5'-terminal group optionally including the substituent  $R^5$ . One of the substituents  $R^2$ ,  $R^{2*}$ ,  $R^3$ , and  $R^{3*}$  is a group  $P^*$  which designates an internucleoside linkage to a preceding monomer, or a 2'/3'-terminal group.

Desirable nucleosides are L-nucleosides such as for example, derived dinucleoside 10 monophosphates. The nucleoside can be comprised of either a beta-D, a beta-L or an alpha-L nucleoside. Desirable nucleosides may be linked as dimers wherein at least one of the nucleosides is a beta-L or alpha-L. B may also designate the pyrimidine bases cytosine, 5-methyl-cytosine, thymine, uracil, or 5-fluorouridine (5-FUdR) other 5-halo compounds, or the purine bases, adenosine, guanosine or inosine.

15 As discussed above, a variety of LNA units may be employed in the monomers and oligomers of the invention including bicyclic and tricyclic DNA or RNA having a 2'-4' or 2'-3' sugar linkages; 2'-O,4'-C-methylene- $\beta$ -D-ribofuranosyl moiety, known to adopt a locked C3'-endo RNA-like furanose conformation. Illustrative modified structures that may be included in oligonucleotides of the invention are shown in Figure 1. Other nucleic acid units that may be 20 included in an oligonucleotide of the invention may comprise 2'-deoxy-2'-fluoro

ribonucleotides; 2'-*O*-methyl ribonucleotides; 2'-*O*-methoxyethyl ribonucleotides; peptide nucleic acids; 5-propynyl pyrimidine ribonucleotides; 7-deazapurine ribonucleotides; 2,6-diaminopurine ribonucleotides; and 2-thio-pyrimidine ribonucleotides, and nucleotides with other sugar groups (e.g. xylose).

- 5 Oligonucleotides containing LNA are readily synthesized by standard phosphoramidite chemistry. The flexibility of the phosphoramidite synthesis approach further facilitates the easy production of LNA oligos carrying all types of standard linkers, fluorophores and reporter groups.

#### 10 Example 18: Selective Binding Complementary (SBC) nucleotides

Selective Binding Complementary (SBC) nucleotides are unable to form stable hybrids with each other, yet are able to form stable, sequence-specific hybrids with complementary unmodified strands of nucleic acids. Thus, the reduced ability of SBC oligonucleotides to form intramolecular hydrogen bond base-pairs between regions of substantially complementary  
15 sequence causes a reduced level of secondary structure. Self-complementarity is an important issue in nucleic acid technologies as reported for DNA, PNA and LNA, and in different biological applications especially in the field of homogeneous assays. LNA:LNA duplexes are the most thermally stable nucleic acid type duplex system known, making the reduction of self-complementarity even more important.

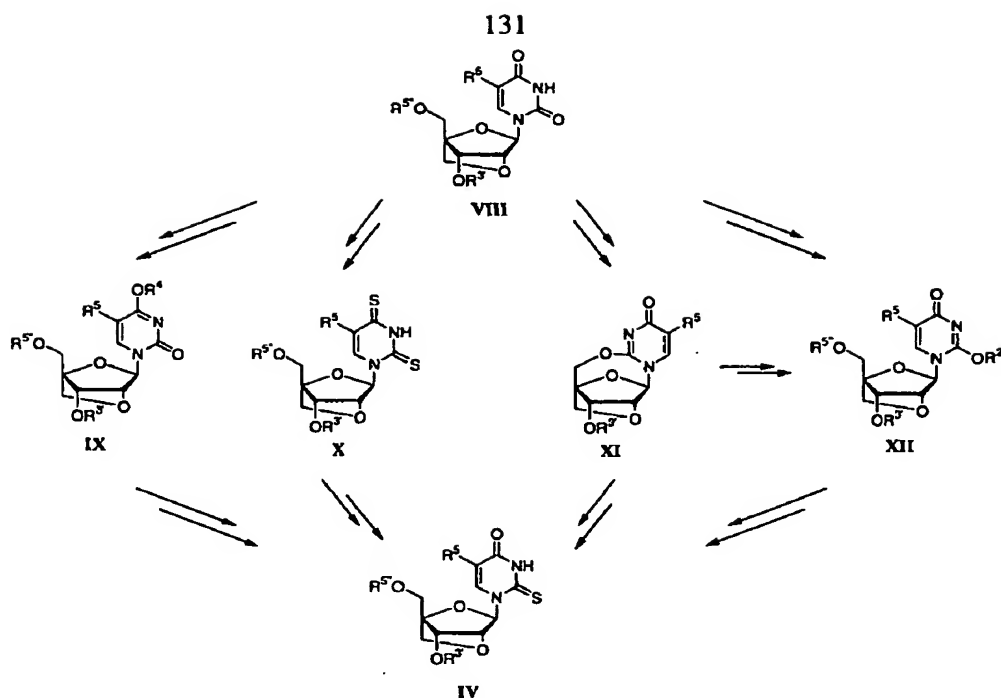
- 20 Exemplary SBC oligonucleotides contain 2-amino-A (D) and <sup>25</sup>T incorporated in the same oligonucleotide as replacements of A and T, respectively. The SBC name refers to the fact that D and <sup>25</sup>T form a destabilised base-pair compared to the A-T base-pair, but D-T and <sup>25</sup>T-A base-pairs are normally more stable than the original A-T base-pair. Exemplary SBC-G nucleotides include inosine or LNA-inosine, and exemplary SBC-C nucleotides include  
25 PyrroloPyr, LNA- PyrroloPyr, <sup>25</sup>C, and LNA-<sup>25</sup>C (Fig. 4). Other exemplary SBC nucleotides are shown in Figs. 2 and 4-8. If desired, SBC nucleotides may be incorporated into the nucleic acids and arrays of the invention, using standard methods.

The systems disclosed herein can provide significant nucleic acid probes for universal hybridization. In particular, universal hybridization can be accomplished with a  
30 conformationally restricted monomer, including a desirable pyrene LNA monomer. Universal hybridization behavior also can be accomplished in an RNA context. Additionally, the binding affinity of probes for universal hybridization can be increased by the introduction of high affinity monomers without compromising the base-pairing selectivity of bases neighboring the universal base.

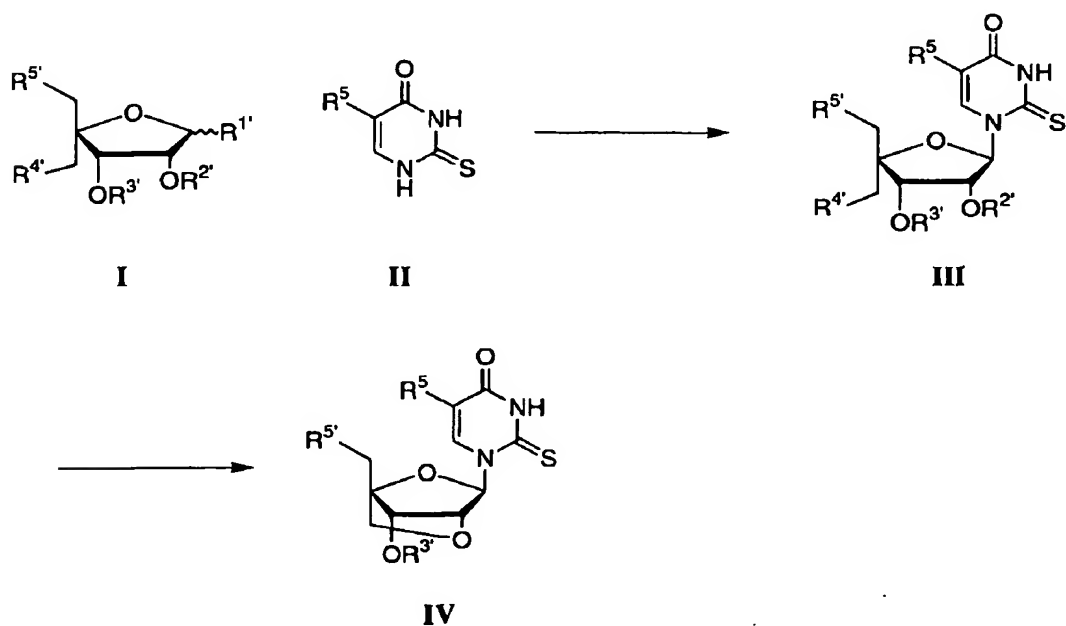
Incorporation of one or more modified nucleobases or nucleosidic bases into an oligonucleotide can provide significant advantages. Among other things, LNA oligonucleotides can often self-hybridize, rather than hybridize to another oligonucleotide. Use of one or more modified bases with the LNA units can modulate the propensity of the oligonucleotide to form double stranded structures with other oligonucleotides containing modified nucleobases including internal duplex formation, thereby inhibiting undesired self-hybridization.

Example 19: Exemplary Methods for Synthesizing LNA-2-thiopyrimidine Nucleosides and Nucleotides

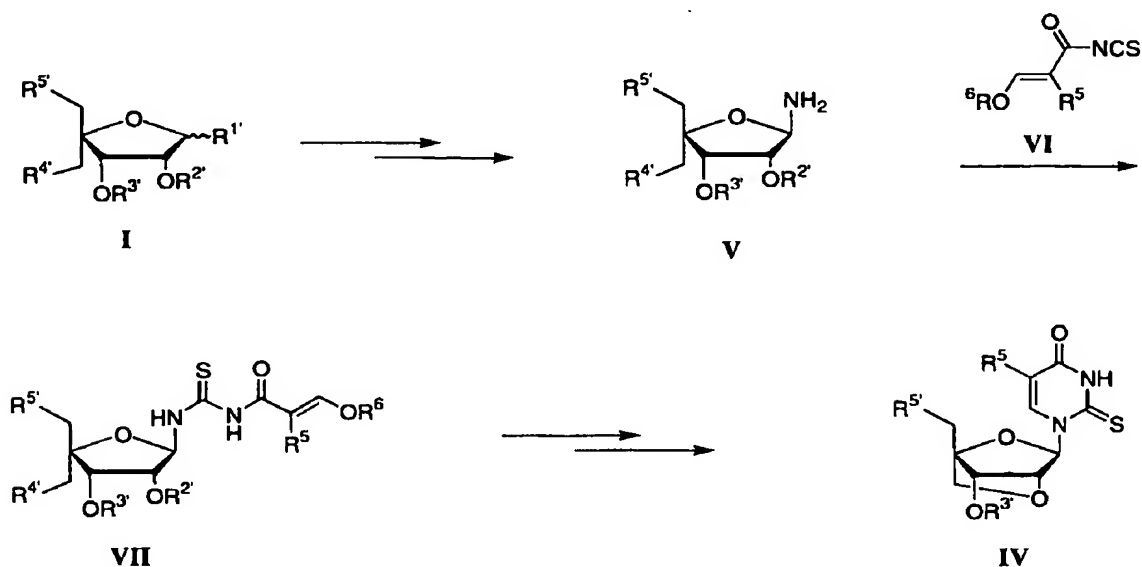
10        2-Thiopyrimidine nucleosides can be prepared in several ways as described below. For example, the 2-thiouridine-nucleosides (IV) can be synthesized from a substituted uridine nucleoside (VIII) as described in the scheme below. By protection of the O4-position (IX) on the nucleobase thionation can be performed, O2 position, which results in the 2-thio-uridine nucleoside (IV). Performing sulphurisation on both O2 and O4 results in 2,4-dithio-uridine  
15 nucleoside (X) which may be transformed into the 2-thio-uridine nucleoside (IV) (Saladino, *et. al.*, Tetrahedron, 1996, 52, 6759). Another way is to generate a cyclic ether (XI) through reaction with the 5' position this product can then be transformed to the 2-thio-uridine nucleoside (IV) or the 2-O-alkyl-uridine nucleoside (XII). The 2-O-alkyl-uridine nucleoside (XII) can also be generated by direct alkylation of the uridine nucleoside (VIII). Treatment of  
20 the 2-O-alkyl-uridine nucleoside (XII) can also be transformed into the 2-thio-uridine nucleoside (Brown *et. al.*, J. Chem. Soc. 1957, 868; Singer, *et. al.*, Proc. Natl. Acad. Sci. USA, 1983, 80, 4884; Rajur and McLaughlin, Tetrahedron Lett., 1992, 33, 6081).



In another method, lewis acid-catalyzed condensation of a properly substituted sugar (**I**) and a substituted 2-thio-uracil (**II**) can result in a substituted 2-thio-uridine nucleoside of the 5 structure (**III**) which by further synthetic manipulations can be transformed into the LNA 2-thiouridine nucleoside (**IV**) (Hamamura et. al., Moffatt, J. Med. Chem., 1972, 15, 1061; Bretner et. al., J. Med. Chem., 1993, 36, 3611).



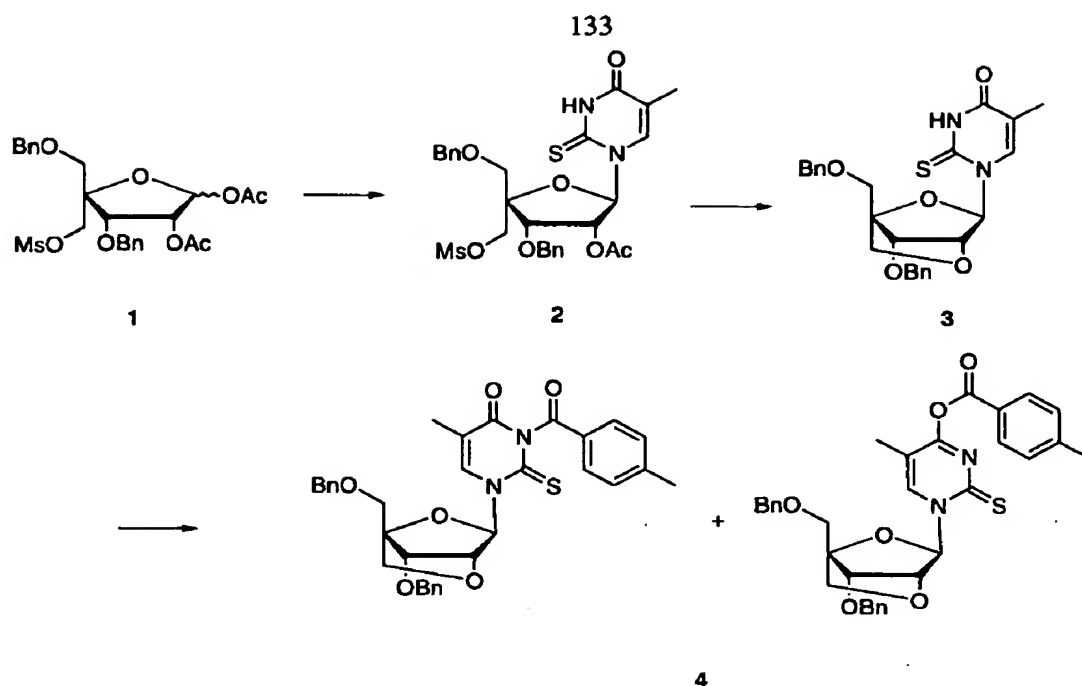
Using a properly substituted amino-sugar (V), a 2-thio-uridine nucleoside can be synthesized through ring-synthesis of the nucleobase by reaction of the amino sugar (V) and an substituted isothiocyanate(VI), yielding the substituted LNA 2-thio-uracil nucleoside (VI) 5 (Shaw and Warren, *J. Chem. Soc.* 1957, 153; Cusack *et al.*, *J. Chem. Soc. Perkin I*, 1973, 1721).



#### 10 Example 20: Exemplary Methods for Synthesizing <sup>2s</sup>T-LNA

Three different strategies for synthesis of <sup>2s</sup>T-LNA are outlined in the Summary of the Invention section. Strategy A involves coupling a glycosyl-donor and a nucleobase, using standard methodology for synthesis of existing LNA monomers. Strategy B involves ring synthesis of the nucleobase. This strategy is desirable because the availability of 1-amino-LNA 15 enables introduction of a variety of new nucleobases. Strategy C includes modification of T-LNA; the easy synthesis of LNA-T diol makes this an attractive pathway.

In a desirable embodiment, <sup>2s</sup>T-LNA is synthesized as illustrated in the scheme below.



In particular, the known coupling sugar 1,2-di-O-acetyl-3, 5 di-O-benzyl, 4-C-mesyloxymethyl,  $\alpha,\beta$ -D-ribofuranose **1** was coupled with the nucleobase 2-thio-thymidine in a Vorbrüggen type reaction. Thus, the nucleobase was silylated and condensed with the sugar using  $\text{SnCl}_4$  as catalyst to promote the reaction affording nucleoside **2**. Mass spectrometry and NMR subsequently identified the isolated product as the desired one. NMR data were compared with published data of a 2-thio-thymidine derivative (Kuimelis and Nambiar, Nucleic Acid Res., 1994, 22, 1429-1436) in order to validate the correct attachment point of the nucleobase.

Subsequently, a base mediated ring-closing reaction afforded the di-benzylated LNA derivative **3** in 77% yield. The signals in the  $^1\text{H}$ -NMR spectrum of the compound appeared as singlets, thus proving that the cyclization had occurred to give the LNA skeleton, in which the 1'-H and 2'-H are perpendicular to each other causing the  $^3J_{1',2'}$  to be 0 Hz. MALDI mass spectrometry was likewise used for the identification of the compound.

The LNA derivative was protected at the nucleobase with the toluoyl protective group to give **4**. This group is well known for the protection of 2-thio-thymidine derivatives, (Kuimelis and Nambiar, Nucleic Acid Res., 1994, 22, 1429-1436). The protection of the nucleobase occurs at both the N-3 and the O-4 position and hence the compound is isolated as a mixture of two compounds. NMR shows that the ratio of the two isomers in the isolated mixture is 2:1.

These methods are described further below.

#### 1-(2-O-acetyl-3-O, 5-O-dibenzyl, 4-C-mesyloxymethyl- $\beta$ -D-ribofuranosyl)-2-thio-thymine (2)

1, 2-di-*O*-acetyl-3, 5 di-*O*-dibenzyl, 4-*C*-mesyloxymethyl,  $\alpha,\beta$ -D-ribofuranose (**1**, 2.0g, 3.83 mmol) and 2-thio-thymine (552mg, 3.89mmol) were co-evaporated with anhydrous acetonitrile (100 ml) and redissolved in anhydrous acetonitrile (80ml), *N,O*-bistrimethylsilylacetamide (1.5, 5.85mmol) was added, and the reaction was stirred at 80°C for 5 one hour. The mixture was cooled to 0°C, SnCl<sub>4</sub> (0.9 ml, 7.66mmol) was added, and the reaction was left to stir for 24 hours. The reaction mixture was diluted with EtOAc and washed with NaHCO<sub>3</sub> and subsequently with water. The organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>) and evaporated to dryness. The product was purified using column chromatography, giving the thio-thymidine derivative **2** (1.1g, 1.82mmol, 40%) as a white foam. R<sub>f</sub> (10% THF/dichloromethane): 0.75.

MALDI-MS: 627 (M+Na) <sup>13</sup>C-NMR (CDCl<sub>3</sub>):  $\delta$ = 174.40, 169.29, 159.89, 136.13, 136.51, 136.05, 128.62, 128.56, 128.41, 128.29, 128.07, 127.89, 127.67, 116.18, 91.41, 86.21, 75.59, 75.31, 74.46, 74.22, 73.61, 69.25, 69.04, 37.52, 20.62, 11.91

15 (1*R*,3*R*,4*R*,7*S*)-7-(benzyloxy)-1-(benzyloxymethyl)-3-(2-thiothymidine)-2,5-dioxabicyclo[2.2.1]heptane (3)

1-(2-*O*-acetyl-3-*O*, 5-*O*-dibenzyl, 4-*C*-mesyloxymethyl- $\beta$ -D-ribofuranosyl)-2-thio-thymine (**2**, 630mg, 1.04mmol) was dissolved in dioxane (15ml) and water (8ml), and aqueous NaOH (2M, 5ml) was added, and the reaction was left to stir at room temperature for one hour. 20 The yellow solution was neutralized with HCl (1M, 6ml) affording a precipitation. The mixture was diluted with dichloromethane and ethyl acetate causing an emulsion. After separation, the aqueous phase extracted with ethyl acetate, and the combined organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>) and evaporated to dryness. The compound was purified by column chromatography (0-2, then 5% THF/dichloromethane), giving the ring closed compound **3** as a white foam 25 (370mg, 0.79mmol, 77%). R<sub>f</sub> (2% MeOH/dichloromethane): 0.23.

MALDI-MS: 488 (M+Na) <sup>13</sup>C-NMR (CDCl<sub>3</sub>):  $\delta$ = 173.14, 160.39, 137.20, 136.63, 136.00, 128.46, 128.34, 128.02, 127.66, 115.52, 90.29, 87.77, 77.39, 75.26, 73.77, 72.07, 71.70, 64.15, 30.17, 12.33

<sup>1</sup>H-NMR (CDCl<sub>3</sub>):  $\delta$ = 9.87 (s, 1H), 7.69 (d, 1.1Hz, 1H), 7.26-7.37 (m, 10H), 6.13 (s, 1H), 4.84 30 (s, 1H), 4.66 (d, J= 11.3 Hz, 1H), 4.61 (s, 2H), 4.52 (d, 11.5Hz, 1H), 4.04 (d, J=7.7Hz, 1H), 3.93 (s, 1H), 3.88 (d, J= 11.0Hz, 1H), 3.82 (d, J= 7.7Hz, 1H), 3.82 (d, J= 10.8 Hz, 1H), 1.59 (d, J= 1.1 Hz, 3H)

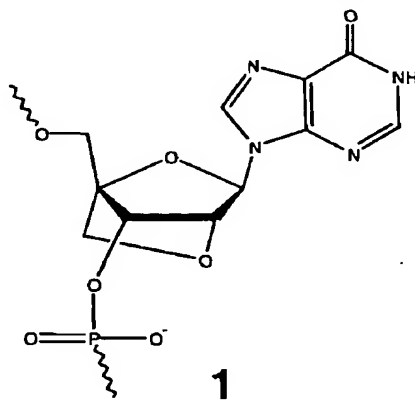
(1*R*,3*R*,4*R*,7*S*)-7-(benzyloxy)-1-(benzyloxymethyl)-3-(2-thio-N3/O4-toluoyl- thymidine)-2,5-dioxabicyclo[2.2.1]heptane (4)

(1*R*,3*R*,4*R*,7*S*)-7-(benzyloxy)-1-(benzyloxymethyl)-3-(2-thiothymidine)-2,5-dioxabicyclo[2.2.1]heptane (**3**, 290mg, 0.62mmol) was dissolved in anhydrous pyridine and 5 diisopropylethylamine (0.2ml, 1.15mmol), toluoyl chloride (0.25ml, 1.89 mmol) was added, and the reaction mixture was stirred at room temperature for three hours. After completion, the reaction mixture was diluted with dichloromethane, and the reaction was quenched by addition of water. The phases were separated, and the organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>) and evaporate to dryness. The residue was co-evaporated with toluene. The product was purified by column 10 chromatography (0-1% MeOH/dichloromethane) to give nucleoside **4** as a white foam (320 mg, 0.55mmol, 89%). R<sub>f</sub> (2%MeOH/dichloromethane): 0.78.

MALDI-MS: 606 (M+Na) <sup>13</sup>C-NMR (CDCl<sub>3</sub>): δ=171.98, 168.30, 160.30, 145.92, 145.82, 137.22, 136.65, 135.98, 130.39, 130.27, 129.85, 129.50, 128.51, 128.41, 128.08, 127.73, 115.11, 90.10, 87.81, 76.01, 75.80, 75.39, 75.01, 73.83, 72.19, 72.09, 71.74, 64.15, 21.75, 15 12.40.

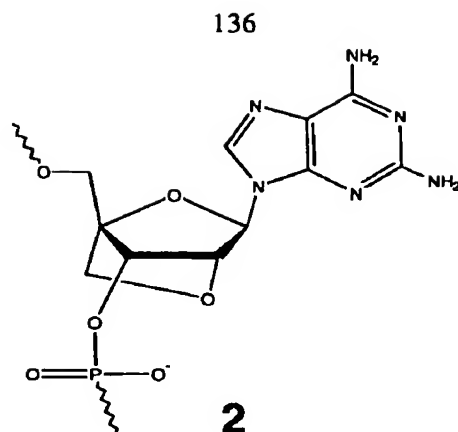
Example 21: Exemplary Methods for Synthesizing LNA-I, LNA-D, and LNA-2AP

2'-*O*, 4'-*C*-methylene linked (LNA) nucleosides containing hypoxanthine (or inosine) (LNA-I), 2,6-diaminopurine (LNA-D), and 2-aminopurine (LNA-2AP) nucleobases were 20 efficiently prepared via convergent syntheses. The nucleosides were converted into phosphoramidite monomers and incorporated into LNA oligonucleotides using an automated phosphoramidite method. The complexing properties of oligonucleotides containing these LNA nucleosides were assessed against perfect and singly mismatch DNA.

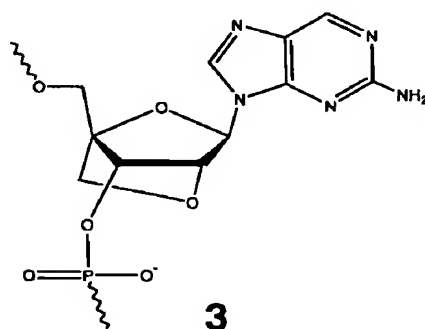


25 LNA-I





LNA-D



LNA-2AP

5

Hypoxanthine, the base found in the nucleotides inosine and deoxyinosine, is considered as a guanine analogue in nucleic acids.

Oligonucleotides containing 2,6-diaminopurine replacements for adenines are expected to bind more strongly to their complementary sequences especially as part of A-type helices 10 due to the potential formation of three hydrogen bonds with thymine or uracil. The reported effect of 2,6-diaminopurine deoxyriboside (D) on the stability of polynucleotide duplexes reaches, on average, about 1.5 °C per modification. Higher stabilization effects for mismatches were observed for D nucleosides involved in formation of duplexes prone to form A-type helices. LNA D and LNA 2'-OMe-D are expected to have increased stabilization and 15 mismatch discrimination. LNA can be used in combination with 2-thio-T for construction of selectively binding complementary oligonucleotides. Taking into consideration the extremely high stability of LNA:LNA duplexes, this approach might be very useful for constructing of LNA containing capture probes and antisense reagents including drugs (Figs. 3 and 10).

2-Aminopurine (2-AP) is a fluorescent nucleobase (emission at 363 nm), which is useful 20 for probing nucleic acids structure and dynamics and for hybridizing with thymine in Watson-Crick geometry. LNA-I, LNA-D, and/or LNA-2AP may be used in the nucleic acids of the

present invention, e.g., to increase the priming efficiency of DNA oligonucleotides in PCR experiments and to construct selectively binding complementary agents.

#### *Synthesis of LNA-I*

The convergent method adopted for preparation of LNA monomers (Koshkin *et al.*, J. Org. Chem. 66:8504, 2001) was successfully applied for syntheses of the modified LNA nucleotides 1-3. The synthetic route to LNA-I phosphoramidite 11 is depicted in the scheme below. The previously described 4-C-branched furanose 4 (Koshkin *et al.*, *supra*) was used as a glycosyl donor in coupling reaction with silylated hypoxanthine by the method of Vorbrüggen *et al.* (Vorbrüggen *et al.*, Chem. Ber. 114:1234, 1981; Vorbrüggen *et al.*, Chem. Ber. 114:1256, 1981; and Vorbrüggen, Acta Biochim. Pol., 43:25, 1996). The reaction resulted in high yield formation of desired  $\beta$ -configured nucleoside derivative 5. However, analogous to the coupling reaction of 4 with protected guanines, the formation of undesired *N*-7 isomer (ratio of *N*-9/*N*-7 = 4:1) was also detected. The mixture of the isomers was used for the ring closing reaction and protected LNA nucleoside 6 was isolated in 68 % yield as a crystalline compound. The correct structure of the isolated isomer was confirmed later by chemical conversion of LNA-I into LNA-A nucleoside (*vide infra*). Deprotection of the 5'-hydroxy group of 6 was accomplished via two-step procedure developed for the syntheses of other LNA nucleosides (Koshkin *et al.*, *supra*). First, 5'-*O*-mesyl group was displaced by sodium benzoate to produce nucleoside 7. The latter was converted into 5'-hydroxy derivative 8 after saponification of the 5'-benzoate. Direct removal of the 3'-*O*-benzyl group from compound 8 was unsuccessful under the conditions tested due to a solubility problem. Therefore, compound 8 was converted to DMT-protected nucleoside 9 prior to catalytic debenzylation of the 3'-*O*-hydroxy group. The phosphoramidite 11 was finally afforded via standard phosphitylation (McBride *et al.*, Tetrahedron Lett. 24:245, 1983; Sinha *et al.*, Tetrahedron Lett. 24:5843, 1983; and Sinha *et al.*, Nucleic Acids Res. 12:4539, 1984) of the nucleoside 10. In order to verify the correct orientation of the glycoside bond (*N*-9 isomer) in synthesized LNA-I nucleoside, compound 7 was successfully converted into the known LNA-A derivative 13 (Koshkin *et al.*, *supra*) (Scheme 2). Thus, a treatment of 7 with phosphoryl chloride according to the procedure reported by Martin (Helv. Chim. Acta 78:486, 1995) resulted in a high yield formation of 6-chloropurine derivative 12. The adenosine derivative 13 was derived from 12 after reaction with ammonia.

#### *Exemplary Analytical Data*

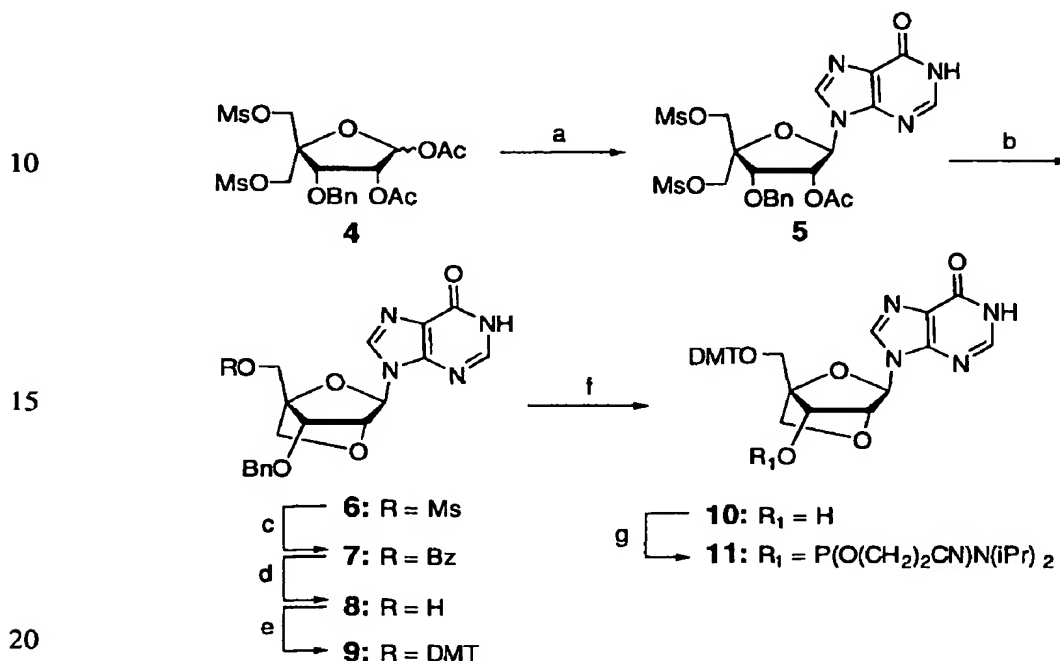
Data for compound 8 includes the following: mp 302-305 °C (dec). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>):  $\delta$  8.16, (s, 1H), 8.06 (s, 1H), 7.30-7.20 (m, 5H), 5.95 (s, 1H), 4.69 (s, 1H), 4.63 (s, 2H), 4.28 (s, 1H), 3.95 (d, *J* = 7.7, 1H), 3.83 (m, 3H). <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>):  $\delta$  156.6, 147.3, 146.1, 137.9,

137.3, 128.3, 127.6, 127.5, 124.5, 88.2, 85.4, 77.0, 72.1, 71.3, 56.7. MALDI-MS  $m/z$ :  $(M+H)^+$ .

Anal. Calcd for  $C_{18}H_{18}N_4O_5 \cdot 5/12 H_2O$ : C, 57.21; H, 5.02; N, 14.82. Found: C, 57.47; H, 4.95; N, 14.17.

Analysis of compound **11** indicated that  $^{31}P$  NMR (DMSO- $d_6$ ):  $\delta$  148.90.

### 5 Scheme for Synthesis of LNA-I<sup>a</sup>



<sup>a</sup> Keys: (a) hypoxanthine, BSA, TMSOTf, 1,2-dichloromethane; 93%; (b) NaOH, THF, EtOH, H<sub>2</sub>O; 69%; (c) NaOBz, DMSO; 76%; (d) NaOH, THF, MeOH, H<sub>2</sub>O; 85%; (e) DMT-Cl, pyridine; 92%; (f) Pd/C, HCO<sub>2</sub>NH<sub>4</sub>; 77%; (g) 2-cyanoethyl-*N,N*-diisopropyl-phosphoramidochloridite, DIPEA, DMF; 75%.

### Exemplary Experimental Conditions

(1*R*,3*R*,4*R*,7*S*)-7-(2-Cyanoethoxy(diisopropylamino)phosphinoxy)-1-(4,4'-dimethoxytrityloxymethyl)-3-(hydroxanthin-9-yl)-dioxabicyclo[2.2.1]heptane (**11**)

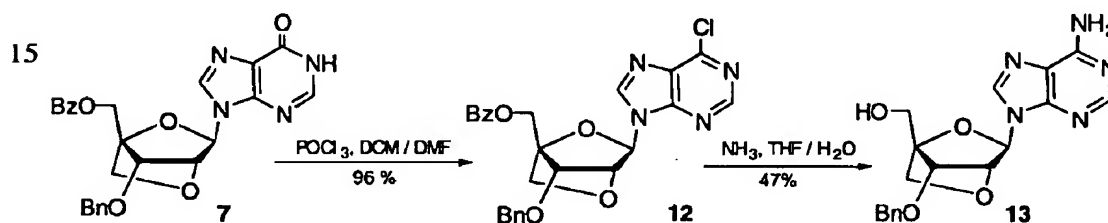
Compound **10** (530 mg, 0.90 mmol, described previously, (see for example, WO 00/56746) was dissolved in anhydrous EtOAc (5 mL) and cooled in an ice-bath. DIPEA (0.47 mL, 2.7 mmol) and (250  $\mu$ L, 1.1 mmol) were added under intensive stirring. Formation of insoluble material was observed, and CH<sub>2</sub>Cl<sub>2</sub> (3 mL) was added to produce a clear solution. More 2-cyanoethyl-*N,N*-diisopropylphosphoramidochloridite (200  $\mu$ L,

0.88 mmol) was added after one hour, and the mixture was stirred overnight. EtOAc (30 mL) was added, the mixture was washed with sat.  $\text{NaHCO}_3$  (2 x 50 mL), brine (50 mL), dried ( $\text{Na}_2\text{SO}_4$ ), and concentrated to a solid residue. Purification by silica gel HPLC (1-5 %MeOH/ $\text{CH}_2\text{Cl}_2$  v/v, containing 0.1% of pyridine) gave compound **11** (495 mg, 75%) as a white solid material.  $^{31}\text{P}$  NMR ( $\text{DMSO}-d_6$ ):  $\delta$  148.90.

#### Synthesis of LNA-D

Taking advantage of a high availability of the natural deoxy- and riboguanosines, a number of effective methods were developed for their conversion into 2,6-diaminopurine (D) nucleosides (Fathi *et al.*, Tetrahedron Lett. 31:319, 1990; Gryaznov *et al.*, Tetrahedron Lett., 10 35:2489, 1994; and Lakshman *et al.*, Org. Lett., 2:927, 2000). However, the production of LNA-G nucleoside is a multi-step synthetic procedure.

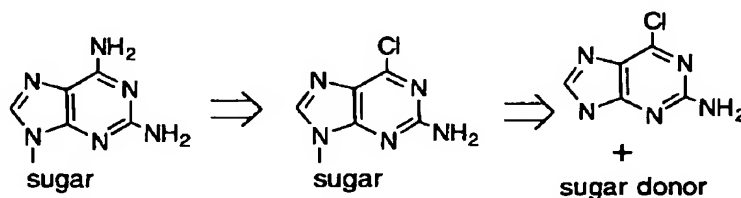
#### Scheme for Synthesis of LNA-G



20

For the synthesis of LNA-D nucleoside, a novel synthesis method was developed that employed a common convergent scheme, related to the strategy used earlier for the synthesis of its anhydrohexitol counterpart (Boudou *et al.*, Nucleic Acids Res. 27:1450, 1999). In particular, a properly protected carbohydrate unit was conjugated with 6-chloro-2-aminopurine **25** to give a stable 6-chloro intermediate derivative (scheme below) which was further converted into desired diaminopurine nucleoside.

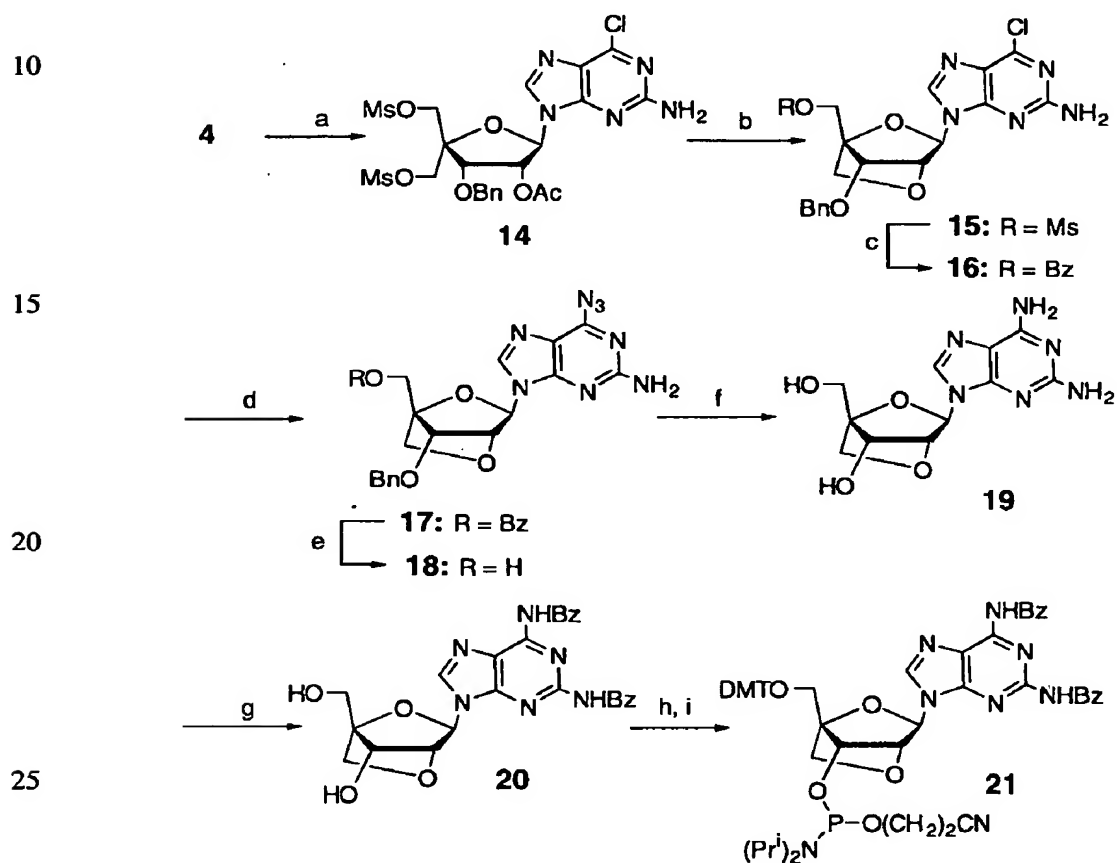
30



35

Thus, it was shown that glycosylation of 2-chloro-6-aminopurine with compound **4** resulted in highly stereoselective formation of the nucleoside derivative **14**. To promote the ring closing 5 reaction, a solution of **14** in aqueous 1,4-dioxane was treated with 10-fold excess of sodium

### Synthesis of LNA-D<sup>a</sup>



<sup>a</sup> Keys: (a) 2-chloro-6-aminopurine, BSA, TMSOTf, 1,2-dichloromethane; 90 %; (b) NaOH, 1,4-dioxane, H<sub>2</sub>O; 87%; (c) NaOBz, DMF; (d) NaN<sub>3</sub>, DMSO; (e) NaOH, EtOH; 79% (three steps); (f) 10% Pd/C, HCO<sub>2</sub>NH<sub>4</sub>, MeOH, H<sub>2</sub>O; 84%; (g) 1. BzCl, pyridine; 2. NaOH, EtOH, pyridine; 62%; (h) DMT-Cl, pyridine; 80%; (i) 2-cyanoethyl-*N,N*-diisopropylphosphoramidochloridite, DIPEA, DMF; 74%.

hydroxide to give bicyclic compound **15** in 87% yield. The standard reaction with sodium benzoate in hot DMF was then successfully applied for displacement of 5'-mesylate of **15**. Notably, this reaction proceeded in very selective manner and no side products originating from the modification of the nucleobase were detected. The desired compound **16** was precipitated from the reaction mixture after addition of water. In order to introduce the 6-amino group into nucleobase structure, intermediate 6-azido derivative **17** was synthesized via reaction of **16** with sodium azide. The nucleoside derivative **18** was isolated as a crystalline compound after saponification of the 5'-benzoate of **17**. Subsequent catalytic hydrogenation of **18** on palladium hydroxide resulted in simultaneous reduction of 6-azido and 3'-benzyl groups to give LNA-D diol **19** after crystallization from water. By the use of peracelation method, 2- and 6-amino groups of **19** were benzoylated at the next step to give the nucleobase protected derivative **20**, which was in the standard way further converted into phosphoramidite monomer **21**.

This phosphoramidite has been produced in a quantity of 0.5 grams.

#### *Exemplary Analytical Data*

**15** Data for compound **19** includes the following:  $^1\text{H}$  NMR (DMSO- $d_6$ ):  $\delta$  7.81 (s, 1H), 6.78 (br s, 2H), 5.91 (br s, 2H), 5.71 (s, 1H), 5.66 (br s, 1H), 5.04 (br s, 1H), 4.31 (s, 1H), 4.20 (s, 1H), 3.90 (d,  $J = 7.7$  Hz, 1H), 3.77 (m, 2H), 3.73 (d,  $J = 7.7$  Hz, 1H).  $^{13}\text{C}$  NMR (DMSO- $d_6$ ):  $\delta$  160.5, 156.2, 150.9, 134.2, 113.4, 88.3, 85.0, 79.3, 71.5, 70.0, 56.8. MALDI-MS  $m/z$ : 295.0 ( $M+H$ ) $^+$ . Anal. Calcd for  $\text{C}_{11}\text{H}_{14}\text{N}_6\text{O}_4 \cdot 1.5 \text{H}_2\text{O}$ : C, 41.12; H, 5.33; N, 26.15. Found: C, 41.24; H, 5.19; N, 25.80.

The  $^{31}\text{P}$  NMR (DMSO- $d_6$ ) spectrum for compound **24** contained signals at  $\delta$  149.19 and 148.98.

Data for compound **23** includes the following: crystallized from MeOH. mp. 227.5-229 °C (dec).  $^1\text{H}$  NMR (DMSO- $d_6$ ):  $\delta$  8.60 (s, 1H), 8.15 (s, 1H), 6.64 (br s, 2H), 5.82 (s, 1H), 5.71 (br s, 1H), 5.04 (br s, 1H), 4.40 (s, 1H), 4.21 (s, 1H), 3.92 (d,  $J = 7.7$  Hz, 1H), 3.79 (m, 2H), 3.75 (d,  $J = 7.7$  Hz, 1H).  $^{13}\text{C}$  NMR (DMSO- $d_6$ ):  $\delta$  160.6, 152.0, 149.4, 139.3, 127.1, 88.6, 84.8, 79.1, 71.6, 70.2, 56.8. MALDI-MS  $m/z$ : 334.7 ( $M+H$ ) $^+$ .

For protected compound **23**, the  $^{31}\text{P}$  NMR (DMSO- $d_6$ ) spectrum has a signal at 148.93 and 148.85.

#### **30 Exemplary Experimental Conditions**

(1S,3R,4R,7S)-3-(2-amino-6-chloropurin-9-yl)-7-benzyloxy-1-methanesulfonoxymethyl-2,5-dioxabicyclo[2.2.1]heptane (15)

To a solution of compound **14** (40 g, 64.5 mmol) in 1,4-dioxane (300 mL) was added 1 M NaOH (350 mL). The mixture was stirred for one hour at 0 °C, neutralized with AcOH (40 mL), and washed with  $\text{CH}_2\text{Cl}_2$  (2 x 200 mL). The combined organic layers were dried ( $\text{Na}_2\text{SO}_4$ )

and concentrated under reduced pressure. The solid residue was purified by silica gel flash chromatography to give compound **15** (27.1 g, 87%) as a white solid material.  $^1\text{H}$  NMR ( $\text{CDCl}_3$ ):  $\delta$  7.84 (s, 1H), 7.32-7.26 (m, 5H), 5.91 (s, 1H), 4.73 (s, 1H), 4.66 (d,  $J = 11.7$  Hz, 1H), 4.61 (d,  $J = 11.7$  Hz, 1H), 4.59 (s, 2H), 4.31 (s, 1H), 4.18 (d,  $J = 8.0$  Hz, 2H), 3.99 (d,  $J = 7.9$  Hz, 1H), 3.05 (s, 3H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta$  158.9, 152.2, 151.4, 139.1, 136.4, 128.4, 128.2, 127.7, 125.3, 86.5, 85.2, 77.2, 76.8, 72.4, 72.1, 64.0, 37.7. MALDI-MS  $m/z$  482.1  $[\text{M}+\text{H}]^+$ .

(1S,3R,4R,7S)-3-(2-amino-6-chloropurin-9-yl)-1-benzoyloxymethyl-7-benzyloxy-2,5-dioxabicyclo[2.2.1]heptane (16)

10 A mixture of sodium benzoate (7.78 g, 54 mmol) and compound **15** 13 g, 27 mmol) was suspended in anhydrous DMF (150 mL) and stirred for two hours at 105 °C. Ice-cold water (500 mL) was added to the solution under intensive stirring. The precipitate was filtered off, washed with water, and dried *in vacuo*. The intermediate product **16** (8 g) was used for ext step without further purification. Analytical sample was additionally purified by silica gel HPLC (0-2% 15 MeOH/ $\text{CH}_2\text{Cl}_2$  v/v).  $^1\text{H}$  NMR ( $\text{CDCl}_3$ )  $\delta$  7.98-7.95 (m, 2H), 7.79 (s, 1H), 7.62-7.58 (m, 1H), 7.48-7.44 (m, 2H), 7.24 (m, 5H), 5.93 (s, 1H), 4.80 (d,  $J = 12.6$  Hz, 1H), 4.77 (s, 1H), 4.67 (d,  $J = 11.9$  Hz, 1H), 4.65 (d,  $J = 12.6$  Hz, 1H), 4.56 (d,  $J = 11.9$  Hz, 1H), 4.27 (d,  $J = 8.0$  Hz, 1H), 4.25 (s, 1H), 4.08 (d,  $J = 7.9$  Hz, 1H).  $^{13}\text{C}$  NMR ( $\text{CDCl}_3$ )  $\delta$  165.7, 158.8, 152.1, 151.3, 138.9, 136.4, 133.4, 129.4, 129.0, 128.5, 128.4, 128.2, 127.6, 125.4, 86.4, 85.7, 77.2, 76.7, 72.5, 72.3, 20 59.5. MALDI-MS  $m/z$  508.0  $[\text{M}+\text{H}]^+$ .

(1S,3R,4R,7S)-3-(2-amino-6-azidopurin-9-yl)-7-benzyloxy-1-hydroxymethyl-2,5-dioxabicyclo[2.2.1]heptane (18)

All the amount of compound **16** from the previous experiment was dissolved in anhydrous DMSO (100 mL) and  $\text{NaN}_3$  (5.4 g, 83 mmol) was added. The mixture was stirred for 25 two hours at 100 °C and cooled to room temperature. Water (400 ml) was added, and the mixture was stirred for 30 minutes at 0 °C (ice-bath) to give a yellowish precipitate **17**. The precipitate was filtered off, washed with water, and dissolved in THF (25 mL). 2M NaOH (30 mL) was then added to the solution, and after 15 minutes of stirring the mixture was neutralized with AcOH (4 mL). The mixture was concentrated to approximately 1/2 of its volume and 30 cooled in an ice-bath. The titel compound was collected by filtration, washed with cold water, and dried *in vacuo*. Yield: 8.8 g (79% from **15**).  $^1\text{H}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  8.53 (br s, 2H), 8.23 (s, 1H), 7.31-7.26 (m, 5H), 6.00 (s, 1H), 5.26 (t,  $J = 5.7$  Hz, 1H), 4.76 (s, 1H), 4.64 (s, 1H), 4.31 (s, 1H), 3.99 (d,  $J = 7.9$  Hz, 1H), 3.88-3.85 (m, 3H).  $^{13}\text{C}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  146.0, 144.0, 143.8, 137.9, 137.0, 128.3, 127.7, 127.6, 112.3, 88.3, 85.6, 77.1, 77.0, 72.2, 71.4, 56.8.

MALDI-MS  $m/z$  384.7  $[M+H]^+$  for 2,6-diaminopurine product, 410.5  $[M+H]^+$ . Anal. Calcd for  $C_{18}H_{18}N_8O_4$ : C, 52.68; H, 4.42; N, 27.30. Found: C, 52.62; H, 4.36; N, 26.94.

(1S,3R,4R,7S)-3-(2,6-Diaminopurin-9-yl)-7-hydroxy-1-hydroxymethyl-2,5-dioxabicyclo[2.2.1]heptane (19)

5 To a suspension of compound 18 (8 g, 19.5 mmol) in MeOH (100 mL) were added  $Pd(OH)_2/C$  (20%, 5.5 g) and  $HCO_2NH_4$  (3g). The mixture was refluxed for 30 minutes and more  $HCO_2NH_4$  (3g) was added. After refluxing for further 30 minutes, the catalyst was filtered off and washed with boiling MeOH/ $H_2O$  (1/1 v/v, 200 mL). The combined filtrates were concentrated to approximately 100 mL and cooled in an ice-bath. The precipitate was filtered  
10 off, washed with ice-cold  $H_2O$  and dried *in vacuo* to give compound 19 (5.4 g, 94 %) as a white solid material.  $^1H$  NMR ( $DMSO-d_6$ ):  $\delta$  7.81 (s, 1H), 6.78 (br s, 2H), 5.91 (br s, 2H), 5.71 (s, 1H), 5.66 (br s, 1H), 5.04 (br s, 1H), 4.31 (s, 1H), 4.20 (s, 1H), 3.90 (d,  $J = 7.7$  Hz, 1H), 3.77 (m, 2H), 3.73 (d,  $J = 7.7$  Hz, 1H).  $^{13}C$  NMR ( $DMSO-d_6$ )  $\delta$  160.5, 156.2, 150.9, 134.2, 113.4, 88.3, 85.0, 79.3, 71.5, 70.0, 56.8. MALDI-MS  $m/z$ : 295.0  $(M+H)^+$ . Anal. Calcd for  
15  $C_{11}H_{14}N_6O_4 \cdot 1.5 H_2O$ : C, 41.12; H, 5.33; N, 26.15. Found: C, 41.24; H, 5.19; N, 25.80.

(1S,3R,4R,7S)-3-(2,6-Di-(N-benzoylamino)purin-9-yl)-7-hydroxy-1-hydroxymethyl-2,5-dioxabicyclo[2.2.1]heptane (20)

A solution of compound 19 (0.5 g, 1.7 mmol) in anhydrous pyridine (20 mL) was cooled in an ice-bath and benzoyl chloride (1.5 mL, 12.9 mmol) was added under intensive  
20 stirring. The mixture was allowed to warm to room temperature and was stirred overnight. Ethanol (20 mL) and 2 M NaOH (20 mL) were added, and the mixture was stirred for an additional hour. EtOAc (75 mL) was added and the solution was washed with water (2 x 50 mL). The combined aqueous layers were washed with  $CH_2Cl_2$  (2 x 50 mL). The combined organic phases were dried ( $Na_2SO_4$ ) and concentrated under reduced pressure to a solid residue.  
25 The residue was suspended in  $Et_2O$  (75 mL, under refluxing for 30 minutes) and cooled in an ice-bath. The product was collected by filtration, washed with cold  $Et_2O$ , and dried *in vacuo* to give compound 20 (530 mg, 62 %) as a slightly yellow solid material.

(1R,3R,4R,7S)-3-(2,6-Di-(N-benzoylamino)purin-9-yl)-1-(4,4'-dimethoxytrityloxymethyl)-7-hydroxy-2,5-dioxabicyclo[2.2.1]heptane (21)

30 Compound 20 (530 mg, 1.06 mmol) was co-evaporated with anhydrous pyridine (2 x 20 mL) and dissolved in anhydrous pyridine (10 mL). DMT-Cl (600 mg, 1.77 mmol) was added, and the solution was stirred overnight at rt. The mixture was diluted with EtOAc (100 mL), washed with saturated  $NaHCO_3$  (100 mL) and brine (50 mL). Organic layer was dried over  $Na_2SO_4$  and concentrated under reduced pressure. Purification by silica gel HPLC (20-100%  
35 EtOAc/hexane v/v, containing 0.1 % of pyridine) gave compound 21 (670 mg, 79%) as a white

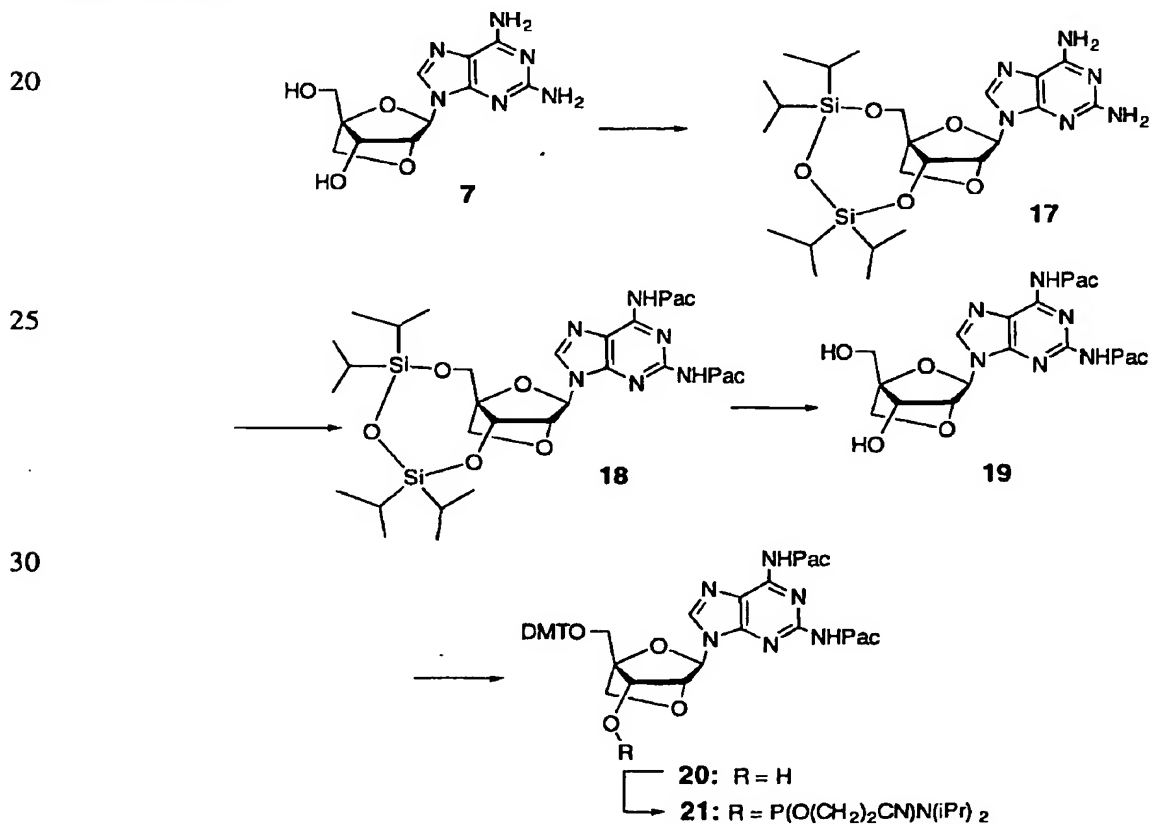


solid material.  $^1\text{H}$  NMR ( $\text{CD}_3\text{OD}$ ):  $\delta$  8.41 (s, 1H), 8.15-8.03 (m, 4H), 7.71-7.22 (m, 15H), 6.92-6.86 (m, 4H), 6.23 (s, 1H), 4.77 (s, 1H), 4.62 (s, 1H), 4.03 (d,  $J = 7.9$  Hz, 1H), 3.99 (d,  $J = 7.9$  Hz, 1H), 3.79 (s, 6H), 3.67 (d,  $J = 10.9$  Hz, 1H), 3.54 (d,  $J = 10.8$  Hz, 1H). MALDI-MS  $m/z$ : 826 ( $\text{M}+\text{Na}$ ) $^+$ . Anal. Calcd for  $\text{C}_{46}\text{H}_{40}\text{N}_6\text{O}_8 \cdot \text{H}_2\text{O}$ : C, 67.14; H, 5.14; N, 10.21. Found: C, 67.24; H, 4.97; N, 10.11.

(1R,3R,4R,7S)-7-(2-Cyanoethoxy(diisopropylamino)phosphinoxy)-3-(2,6-di-(N-benzoylamino)purin-9-yl)-1-(4,4'-dimethoxytrityloxymethyl)-2,5-dioxabicyclo[2.2.1]heptane (21). To a stirred solution of compound **20** (640 mg, 0.8 mmol) in anhydrous DMF (5 mL) were added DIPEA (420  $\mu\text{L}$ , 2.4 mmol) and 2-cyanoethyl-*N,N*-diisopropylphosphoramidochloridite (300  $\mu\text{L}$ , 1.2 mmol). The mixture was stirred for 1.5 hours at room temperature, diluted with EtOAc (100 mL), and washed with saturated  $\text{NaHCO}_3$  (2 x 100 mL) and brine (50 mL). Organic layer was dried ( $\text{Na}_2\text{SO}_4$ ) and concentrated under reduced pressure to give a yellow solid residue. Purification by silica gel HPLC (20-100 % EtOAc/hexene containing 0.1 % of pyridine) gave compound **21** (590 mg, 74%) as a white solid material.  $^{31}\text{P}$  NMR ( $\text{DMSO}-d_6$ )  $\delta$  149.19, 148.98.

#### Synthesis of Pac-protected LNA-D amidite

The following scheme illustrates a method for synthesizing a Pac-protected version of LNA-D amidite.



Compound 17

Compound **7** (1g, 3.39 mmol) was co-evaporated with anhydrous DMF (2 x 10 mL) and dissolved in DMF (10 mL). Imidazole (0.69 g, 10.17 mmol) and 1,3-dichloro-1,1,3,3,5-tetraisopropylidisiloxane (1.4 mL, 4.37 mmol) were added, and the mixture was stirred overnight. H<sub>2</sub>O (100 mL) was added under intensive stirring to precipitate nucleoside material. The precipitate was filtered off, washed with H<sub>2</sub>O, and dried *in vacuo*. Crystallization from ethanol gave compound **17** (1.15 g, 63%) as a white solid material. MALDI-MS: *m/z* 537.3 (M+H)<sup>+</sup>.

10

Compound 18

To a solution of compound **17** (1.15 g, 2.14 mmol) in anhydrous pyridine (5 mL) was added phenoxyacetic anhydride (2 g, 7.0 mmol) and the mixture was stirred for four hours. EtOAc (100 mL) was added, and the solution was washed with sat. NaHCO<sub>3</sub> (2 x 100 mL), 15 brine (50 mL), dried (Na<sub>2</sub>SO<sub>4</sub>), and concentrated to a solid residue. Purification by silica gel HPLC (50-100% v/v EtOAc/hexane) gave compound **18** (1.65 g, 95%) as a white solid material. MALDI-MS: *m/z* 827.3 (M+Na)<sup>+</sup>.

(1S,3R,4R,7S)-3-(2,6-Di-(N-phenoxyacetyl-amino)purin-9-yl)-7-hydroxy-1-hydroxymethyl-2,5-dioxabicyclo[2.2.1]heptane (19)

20 To a solution of compound **18** (0.96 g, 1.19 mmol) in anhydrous THF (10 mL) was added Et<sub>3</sub>N·3HF (0.2 mL) and the mixture was stirred overnight at room temperature. The formed precipitate was collected by filtration and washed with THF (5 mL) and pentane (5 mL) to give after drying compound **19** (650 mg, 97%) as a white solid material. MALDI-MS: *m/z* 563.0 (M+H)<sup>+</sup>.

25 (1R,3R,4R,7S)-3-(2,6-Di-(N-phenoxyacetyl-amino)-purin-9-yl)-1-(4,4'-dimethoxytrityloxymethyl)-7-hydroxy-2,5-dioxabicyclo[2.2.1]heptane (20)

To a solution of compound **19** (650 mg, 1.15 mmol) was added DMT-Cl (500 mg, 1.48 mmol). The mixture was stirred for five hours, diluted with EtOAc (100 mL), and washed with sat. NaHCO<sub>3</sub> (2 x 100 mL). The organic layer was dried and concentrated to a solid residue. 30 Crystallization from EtOAc gave compound **20** (810 mg, 81%) as a white solid material.

(1R,3R,4R,7S)-7-(2-Cyanoethoxy(diisopropylamino)phosphinoxy)-3-(2,6-di-(N-phenoxyacetyl-amino)-purin-9-yl)-1-(4,4'-dimethoxytrityloxymethyl)-2,5-dioxabicyclo[2.2.1]heptane (21)

To a solution of compound **20** (800 mg, 0.92 mmol) in anhydrous DMF (10 mL) were 35 added 0.75 M solution of DCI in EtOAc (0.7 mL) and 2-cyanoethyl

tetraisopropylphosphorodiamidite (0.32 mL, 1.01 mmol). The mixture was stirred at room temperature overnight and EtOAc (75 mL) was added. The resulting solution was washed with sat.  $\text{NaHCO}_3$  and brine, dried and concentrated to a solid residue. Purification by silica gel HPLC (30-100% v/v EtOAc/hexane, containing 0.1% of pyridine) gave phosphoramidite **21** (550 mg, 56%) as a white solid material.

$^{31}\text{P}$  NMR ( $\text{DMSO}-d_6$ ):  $\delta$  149.08, 148.8.

#### Synthesis of LNA-2AP

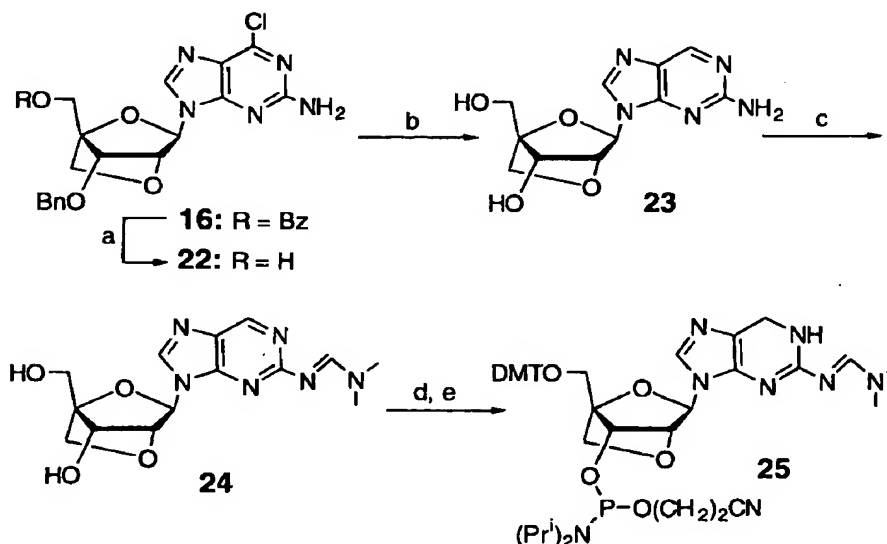
The intermediate derivative **16** was also used for the synthesis of LNA-2AP nucleoside. First, the 5'-*O*-benzoyl group of **16** was hydrolyzed by aqueous sodium hydroxide to give the nucleoside derivative **22** in 72% yield. The conditions of catalytic transfer hydrogenation usually used for removal of the 3'-*O*-benzyl group turned out to be suitable for complete dechlorination of the nucleobase of **22**. Thus, totally deprotected LNA-2AP nucleoside **23** was afforded in high yield after refluxing of the methanolic solution of **22** in the presence of palladium hydroxide and ammonium formate. The 2-amine of **23** was selectively protected with an amidine group after treatment with *N,N*-dimethylformamide dimethyl acetal. The resulting diol **24** was then 5'-*O*-DMT protected and 3'-*O*-phosphitylated to yield the desired phosphoramidite LNA-2AP monomer **25** (McBride *et al.*, J. Am. Chem. Soc. 108:2040, 1986).

#### Synthesis of LNA-2AP<sup>a</sup>

20

25

30



<sup>a</sup> Keys: (a) NaOH, 1,4-dioxane, H<sub>2</sub>O; 72%; (b) 20% Pd(OH)<sub>2</sub>/C, HCO<sub>2</sub>NH<sub>4</sub>, MeOH, H<sub>2</sub>O; 89%; (c) *N,N*-dimethylformamide dimethyl acetal, DMF; (d) DMT-Cl, pyridine; 87% (two steps); (e) 2-cyanoethyl-*N,N*-diisopropylphosphoramidochloridite, DIPEA, DMF; 64%.

5

*Exemplary Experimental Conditions*

(1*S*,3*R*,4*R*,7*S*)-3-(2-amino-6-chloropurin-9-yl)-7-benzyloxy-1-hydroxymethyl-2,5-dioxabicyclo[2.2.1]heptane (22)

To a solution of compound **16** (3 g, 5.92 mmol) in 1,4-dioxane (20 mL) was added 2 M NaOH (20 mL) and the mixture was stirred for one hour. AcOH (3 mL) was added, and the solvents were removed under reduced pressure. The solid residue was re-dissolved in 20% MeOH/EtAc (50 mL), washed with NaHCO<sub>3</sub> (2 x 50 mL), dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated to a solid residue. The residue was purified by silica gel column chromatography (1-2% MeOH/EtAc v/v) to give compound **22** (1.72 g, 72%) as a white solid material.

15 (1*S*,3*R*,4*R*,7*S*)-3-(2-aminopurin-9-yl)-7-hydroxy-1-hydroxymethyl-2,5-dioxabicyclo[2.2.1]heptane (23)

To a solution of compound **22** (0.72 g, 1.79 mmol) in MeOH/dioxane (1/1 v/v) were added Pd(OH)<sub>2</sub>/C (20%, 0.5 g) and HCO<sub>2</sub>NH<sub>4</sub> (1.5 g, 23.8 mmol). The mixture was stirred under refluxing for 30 minutes and cooled to room temperature. The catalyst was filtered off and washed with MeOH. The combined filtrates were concentrated under reduced pressure to yield compound **23** (0.44 g, 89 %) as a white solid material. Analytical sample was crystallized from MeOH. mp. 227.5-229 °C (dec). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>): δ 8.60 (s, 1H), 8.15 (s, 1H), 6.64 (br s, 2H), 5.82 (s, 1H), 5.71 (br s, 1H), 5.04 (br s, 1H), 4.40 (s, 1H), 4.21 (s, 1H), 3.92 (d, *J* = 7.7 Hz, 1H), 3.79 (m, 2H), 3.75 (d, *J* = 7.7 Hz, 1H). <sup>13</sup>C NMR (DMSO-*d*<sub>6</sub>): δ 160.6, 152.0, 20 and 149.4, 139.3, 127.1, 88.6, 84.8, 79.1, 71.6, 70.2, 56.8.

(1*R*,3*R*,4*R*,7*S*)-1-(4,4'-dimethoxytrityloxymethyl)-3-(2-*N*-(dimethylaminomethylidene)aminopurin-9-yl)-7-hydroxy-2,5-dioxabicyclo[2.2.1]heptane (5' DMT protected version of 24)

Compound **23** (0.4 g, 1.43 mmol) was co-evaporated with anhydrous DMF (10 mL) and dissolved in DMF (15 mL). *N,N*-Dimethylformamide dimethylacetal (0.8 mL) was added and the solution was stirred for three days at room temperature. Water (5 mL) was added, and the solvents were removed under reduced pressure. The solid residue was co-evaporated with anhydrous pyridine (2 x 10 mL) and dissolved in anhydrous pyridine (5 mL). DMT-Cl (0.7 g, 2.1 mmol) was added, the solution was stirred for four hours, diluted with EtOAc (50 mL), and washed with NaHCO<sub>3</sub> (2 x 50 mL) and brine (50 mL). Organic layer was dried (Na<sub>2</sub>SO<sub>4</sub>) and

concentrated to a yellow solid residue. Purification by silica gel HPLC (1-6% MeOH/CH<sub>2</sub>Cl<sub>2</sub> v/v, containing 0.1% of pyridine) gave the 5' DMT protected version of compound **24** (0.87 g, 87%) as a white solid material.

(1*R*,3*R*,4*R*,7*S*)-7-(2-Cyanoethoxy(diisopropylamino)phosphinoxy)-1-(4,4'-

5 dimethoxytrityloxymethyl)-3-(2-*N*-(dimethylaminomethylidene)aminopurin-9-yl)-2,5-dioxabicyclo[2.2.1]heptane (25)

The 5' DMT protected version of compound **24** (0.5 g, 0.79 mmol) was dissolved in anhydrous DMF (10 mL) and DIPEA (350 µL) and 2-cyanoethyl-*N,N*-diisopropylphosphoramidochloridite (250 µL) were added. The mixture was stirred for one hour, diluted with EtOAc (50 mL), washed with saturated NaHCO<sub>3</sub> (2 x 100 mL) and brine (50 mL), dried (Na<sub>2</sub>SO<sub>4</sub>), and concentrated to a solid residue. Purification by silica gel HPLC (0-3% MeOH/CH<sub>2</sub>Cl<sub>2</sub> v/v, containing 0.1% of pyridine) gave compound **25** (0.42 g, 64%) as a white solid material. <sup>31</sup>P NMR (DMSO-*d*<sub>6</sub>) δ 148.93, 148.85.

### 15 Synthesis of Oligomers

Along with previously described LNA phosphoramidites (Koshkin *et al.*, *supra*; and Pedersen *et al.*, Synthesis p. 802, 2002), the phosphoramidite monomers **11**, **21**, and **25** were successfully applied for automated oligonucleotide synthesis (Caruthers, Acc. Chem. Res. 24:278, 1991) to produce the LNA oligomers depicted in Table 9. Oligonucleotide syntheses were performed on a 0.2 µmol scale using an Expedite synthesizer (Applied Biosystems) with the recommended commercial reagents. Standard protocols for DNA synthesis were used, except that the coupling time was extended to 5 minutes and the oxidation time was extended to 30 second cycles. Deprotection of the oligonucleotides were performed by treatment with concentrated ammonium hydroxide for five hours at 60 °C. After that, the LNA-D containing oligonucleotides were additionally treated with AMA (concentrated ammonium hydroxide / 40% aqueous MeNH<sub>2</sub>; 1/1 v/v) for one hour at 60°C. All the synthesized oligonucleotides were purified by RP-HPLC, and their structures were verified by MALDI-TOF mass spectra.

The complexing properties of oligonucleotides containing new LNA monomers **1-3** were assessed. Comparative binding data from an 8-mer LNA sequence is shown in Table 9 as the melting temperatures against complementary single stranded DNA. An exemplary sequence for this comparison is GACATAGG, which is the central part of a capture probe used for SNP detection in GlucIVS7-7asA (A:a mismatch position). The thermal stabilities of reference DNA duplexes (entries 1-7, Table 9) can be directly compared with their LNA counterparts (entries 8-14). The hybridizing ability of all LNA 8-mers is superior to that of isosequential DNA oligonucleotides. The average melting temperatures of DNA and LNA 8-

mers against complementary DNAs typically differ by about 40 °C. The replacement of one internal LNA-A nucleotide by LNA-D resulted in the further stabilization of the complementary duplex (i.e., compare entries 8 and 11) by 6.2 °C. Interestingly, the analogous replacement made in an DNA octamer destabilized the corresponding duplex by 0.5 °C (i.e., entries 1 and 5 4). D-nucleosides may facilitate a B to A helix transition, because the A-type structure of an LNA:DNA duplex is more suitable for effective D:t pairing. This stabilizing effect is expected to be even more pronounced for LNA:RNA duplexes, which can be very useful for construction of antisense or other gene-silencing reagents. The mismatch discrimination ability of the D-nucleoside was also studied (entry 11). In comparison to LNA-A (entry 8) D-nucleoside 10 demonstrated remarkable increased mismatch discrimination against DNA-g nucleoside.

Table 9. Melting temperatures ( $T_m$ ) of the complementary DNA-DNA and LNA-DNA duplexes.<sup>a</sup> Modified monomers (LNA are in CAPITALS): I = inosine; D = 2,6-diaminopurine; X = 2-aminopurine.

15

Entry	Oligonucleotide structure	$T_m$ ( $\pm 0.5$ °C) of the duplexes with complementary deoxynucleotide			
		3'-ctgtatcc	3'-ctgaatcc	3'-ctggatcc	3'-ctgeatcc
1	5'-gacatagg	23.8	<10	<10	<10
2	5'-gacttagg	<10	22.6	<10	<10
3	5'-gacgtagg	<10	<10	<10	25.0
4	5'-gacdtagg	23.3	<10	<10	<10
5	5'-gdcddidgg	33.4	<10	<10	17.7
6	5'-gacitagg	<10	<10	<10	20.9
7	5'-gacxtagg	<10	<10	<10	<10
8	5'-GACATAGG	61.6	38.2	43.4	40.6
9	5'-GACTTAGG	28.0	60.7	36.4	23.5
10	5'-GACGTAGG	55.0	32 <sup>b</sup>	41 <sup>b</sup>	70.9
11	5'-GACDTAGG	67.8	42.2	41.4	52.4
12	5'-GDCDTDGG	78.3	55.9	54.7	63.8
13	5'-GACITAGG	53.1	48.2	43.0	59.9
14	5'-GACXTAGG	60.8	45.5	44.0	53.9

<sup>a</sup> The melting temperatures ( $T_m$  values) were obtained as a maxima of the first derivative of the corresponding melting curves (optical density at 260 nm versus temperature). Concentration of the duplexes: 2.5  $\mu$ M. Buffer: 0.1 M NaCl; 10 mM Na-phosphate (pH 7.0); 1 mM EDTA.

<sup>b</sup> Low cooperativity of transitions (accuracy  $\pm 1$  °C).

5

Table 10. The mismatch discrimination effect of the chimeric LNA-DNA 12-mers containing LNA-A or LNA-D nucleosides against the point of mutation

The structure of LNA-DNA oligonucleotide	Tm (± 0.5 °C) of the complementary duplexes with DNA oligonucleotides (Δ Tm between singly mismatched and perfect duplexes)	
HNFas128A-2		
	caacatcccaca	caacaaccaca
tGtggGATGtgg	61.0	45.9 (-15.1)
tGtggGDTGtgg	65.5	49.7 (-15.8)
Gluc53as-A		
	aagagtccagtg	aagaggccagtg
cAmCtgGAmCtctt	61.5	50.6 (-10.9)
cAmCtgGDmCtctt	65.3	45.4 (-19.9)

10 <sup>a</sup> Concentration of duplexes: 2  $\mu$ M; Buffer: see Table 9.

Table 11. Melting temperatures of the LNA and DNA duplexes (LNAs are CAPITALIZED) containing 2-thio-deoxythymidine (s) and diaminopurineriboside (d). See Table 9 for experimental conditions.

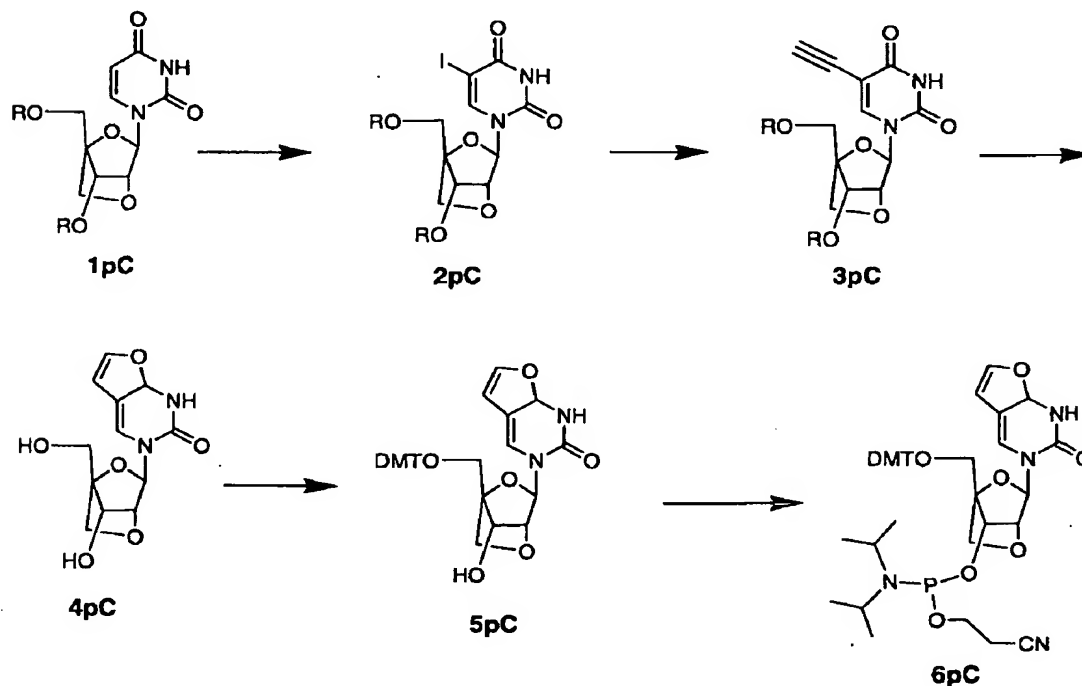
oligo structure	$T_m$ ( $\pm 0.5$ °C) of complementary duplexes with				
	3'-ctgtatcc	3'-ctgsatcc	3'-CTGsATCC	3'-CTGtATCC	3'-CTGTATCC
5'-gacatagg	23.8	27	54.4	49.4	54.6
5'-gacdtagg	23.3	<6	45.4	55.2	60.5
5'-GACATAGG	61.6	64.6	87	88	88
5'-GACDTAGG	67.8	59.4	80	>90	>90

15

\*  $T_m$  values in the shaded cells were measured in low salt buffers (1 mM Na-phosphate, pH 7.0). Low cooperativity of the transitions was observed (accuracy  $\pm 1.5^\circ\text{C}$ )

#### Example 22: Exemplary Methods for Synthesizing LNA-PyrroloPyr-SBC-C

5 The furanopyrimidine phosphoramidite **6pC** used for incorporation of the pyrroloC analogue can be synthesized from LNA-U through a series of reactions as illustrated below and in Fig. 14. Starting from LNA-U **1pC** iodine can be introduced on the 5 position on the nucleobase (Chang and Welch, J. Med. Chem. 1963, 6, 428). This compound can be used in a Sonogashira type palladium coupling reaction (Sonogashira, Tohda and Hagihara, Tetrahedron 10 Lett. 1975, 4467) resulting in the 5-ethynyl-LNA-U **3pC**. The 5-ethynyl-LNA-U **3pC** can be transformed to the furanopyrimidine LNA analogue **4pC** when reacted with CuI, and then transformed into the DMT-protected phosphoramidite **5pC** (Woo, Meyer, and Gamper, Nucleic Acids Res., 1996, 24, 2470). LNA-PyrroloPyr-SBC-C is formed when **6pC** or an oligonucleotide containing **6pC** is deprotected with ammonia.



15

#### Example 23: Exemplary Modified Bases such as Universal Bases

Desirable modified bases are covalently linked to the 1'-position of a furanosyl ring, particularly to the 1'-position of a 2',4'-linked furanosyl ring, especially to the 1'-position of a 2'-O,4'-C-methylene-beta-D-ribofuranosyl ring.



As discussed above, other desirable modified bases contain one or more carbon alicyclic or carbocyclic aryl units, i.e. non-aromatic or aromatic cyclic units that contain only carbon atoms as ring members. Modified bases that contain carbocyclic aryl groups are generally desirable, particularly a moiety that contains multiple linked aromatic groups, particularly 5 groups that contain fused rings. That is, optionally substituted polynuclear aromatic groups are especially desirable such as optionally substituted naphthyl, optionally substituted anthracenyl, optionally substituted phenanthrenyl, optionally substituted pyrenyl, optionally substituted chrysenyl, optionally substituted benzanthracenyl, optionally substituted dibenzanthracenyl, optionally substituted benzopyrenyl, with substituted or unsubstituted pyrenyl being particularly 10 desirable.

Without being bound by any theory, it is believed that such carbon alicyclic and/or carbocyclic aryl modified bases can increase hydrophobic interaction with neighboring bases of an oligonucleotide. Those interactions can enhance the stability of a hybridized oligo pair, without necessity of interactions between bases of the distinct oligos of the hybridized pair.

15 Again without being bound by any theory, it is further believed that such hydrophobic interactions can be particularly favored by platelike stacking of neighboring bases, i.e. intercalation. Such intercalation will be promoted if the base comprises a moiety with a relatively planar extended structure, such as provided by an aromatic group, particularly a carbocyclic aryl group having multiple fused rings. This is indicated by the increases in  $T_m$  20 values exhibited by oligos having LNA units with pyrenyl nucleobases relative to comparable oligos having LNA units with naphthyl nucleobases.

Modified bases that contain one or more heteroalicyclic or heteroaromatic groups also are suitable for use in LNA units, particularly such non-aromatic and aromatic groups that contains one or more N, O or S atoms as ring members, particularly at least one sulfur atom, 25 and from 5 to about 8 ring members. Also desirable is a nucleobase that contains two or more fused rings, where at least one of the rings is a heteroalicyclic or heteroaromatic group containing 1, 2, or 3 N, O, or S atoms as ring members.

In general, desirable are modified bases that contain 2, 3, 4, 5, 6, 7 or 8 fused rings, which may be carbon alicyclic, heteroalicyclic, carbocyclic aryl and/or heteroaromatic; more 30 desirably modified bases that contain 3, 4, 5, or 6 fused rings, which may be carbon alicyclic, heteroalicyclic, carbocyclic aryl and/or heteroaromatic, and desirably the fused rings are each aromatic, particularly carbocyclic aryl.

In some embodiments, the base is not an optionally substituted oxazole, optionally substituted imidazole, or optionally substituted isoxazole modified base.

Other suitable modified bases for use in LNA units in accordance with the invention include optionally substituted pyridyloxazole, optionally substituted pyrenylmethylglycerol, optionally substituted pyrrole, optionally substituted diazole and optionally substituted triazole groups.

5 Desirable modified bases of the present invention when incorporated into an oligonucleotide containing all LNA units or a mixture of LNA and DNA or RNA units will exhibit substantially constant  $T_m$  values upon hybridization with a complementary oligonucleotide, irrespective of the bases present on the complementary oligonucleotide.

In some embodiments, one or more of the common RNA or commonly used derivatives  
10 thereof, such as 2'-O-methyl, 2'-fluoro, 2'-allyl, and 2'-O-methoxyethoxy derivatives are combined with at least one nucleotide with a universal base to generate an oligonucleotide having between five to 100 nucleotides.

Modified nucleic acid compounds may comprise a variety of nucleic acid units e.g. nucleoside and/or nucleotide units. As discussed above, an LNA nucleic acid unit has a carbon  
15 or hetero alicyclic ring with four to six ring members, e.g. a furanose ring, or other alicyclic ring structures such as a cyclopentyl, cycloheptyl, tetrahydropyranyl, oxepanyl, tetrahydrothiophenyl, pyrrolidinyl, thianyl, thiepanyl, piperidinyl, and the like.

In an aspect of the invention, at least one ring atom of the carbon or hetero alicyclic group is taken to form a further cyclic linkage to thereby provide a multi-cyclic group. The  
20 cyclic linkage may include one or more, typically two atoms, of the carbon or hetero alicyclic group. The cyclic linkage also may include one or more atoms that are substituents, but not ring members, of the carbon or hetero alicyclic group.

Unless indicated otherwise, an alicyclic group as referred to herein is inclusive of group having all carbon ring members as well as groups having one or more hetero atom (e.g. N, O, S  
25 or Se) ring members. The disclosure of the group as a "carbon or hetero alicyclic group" further indicates that the alicyclic group may contain all carbon ring members (i.e. a carbon alicyclic) or may contain one or more hetero atom ring members (i.e. a hetero alicyclic). Alicyclic groups are understood not to be aromatic, and typically are fully saturated within the ring (i.e. no endocyclic multiple bonds).

30 Desirably, the alicyclic ring is a hetero alicyclic, i.e. the alicyclic group has one or more hetero atoms ring members, typically one or two hetero atom ring members such as O, N, S or Se, with oxygen being often desirable.

The one or more cyclic linkages of an alicyclic group may be comprised completely of carbon atoms, or generally more desirable, one or more hetero atoms such as O, S, N or Se,

desirably oxygen for at least some embodiments. The cyclic linkage will typically contain one or two or three hetero atoms, more typically one or two hetero atoms in a single cyclic linkage.

The one or more cyclic linkages of a nucleic acid compound of the invention can have a number of alternative configurations and/or configurations. For instance, cyclic linkages of 5 nucleic acid compounds of the invention will include at least one alicyclic ring atom. The cyclic linkage may be disubstituted to a single alicyclic atom, or two adjacent or non-adjacent alicyclic ring atoms may be included in a cyclic linkage. Still further, a cyclic linkage may include a single alicyclic ring atom, and a further atom that is a substituent but not a ring member of the alicyclic group.

10 For instance, as discussed above, if the alicyclic group is a furanosyl-type ring, desirable cyclic linkages include the following: C-1', C-2'; C-2', C-3'; C-2', C-4'; or a C-2', C-5' linkage.

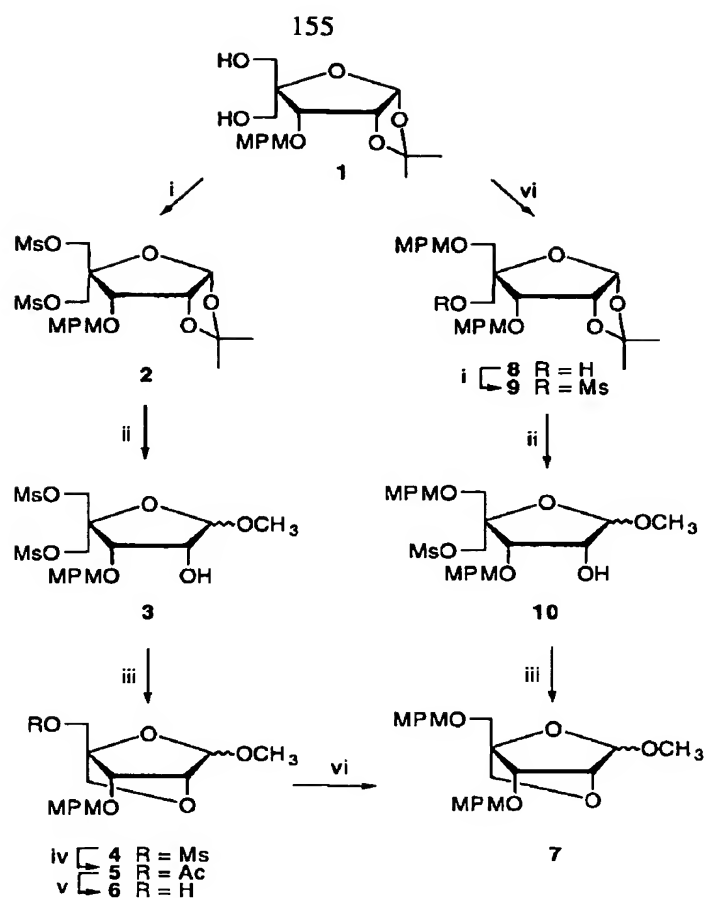
A cyclic linkage will typically comprise, in addition to the one or more alicyclic group ring atoms, 2 to 6 atoms in addition to the alicyclic ring members, more typically 3 or 4 atoms in addition to the alicyclic ring member(s).

15 The alicyclic group atoms that are incorporated into a cyclic linkage are typically carbon atoms, but hetero atoms such as nitrogen of the alicyclic group also may be incorporated into a cyclic linkage.

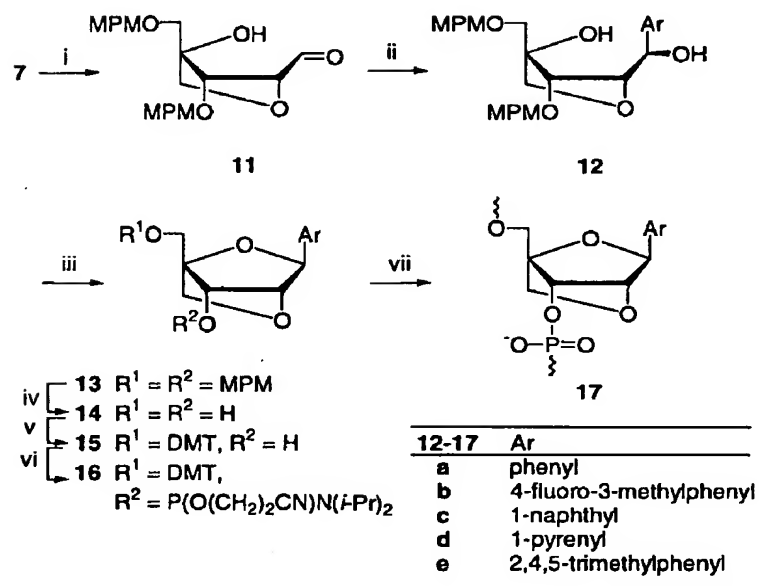
Specifically desirable modified nucleic acids for use oligonucleotides of the invention include locked nucleic acids as disclosed in WO99/14226 (which include bicyclic and tricyclic 20 DNA or RNA having a 2'-4' or 2'-3' sugar linkages); 2'-deoxy-2'-fluoro ribonucleotides; 2'-*O*-methyl ribonucleotides; 2'-*O*-methoxyethyl ribonucleotides; peptide nucleic acids; 5-propynyl pyrimidine ribonucleotides; 7-deazapurine ribonucleotides; 2,6-diaminopurine ribonucleotides; and 2-thio-pyrimidine ribonucleotides.

LNA units as disclosed in WO 99/14226 are in general particularly desirable modified 25 nucleic acids for incorporation into an oligonucleotide of the invention. Additionally, the nucleic acids may be modified at either the 3' and/or 5' end by any type of modification known in the art. For example, either or both ends may be capped with a protecting group, attached to a flexible linking group, attached to a reactive group to aid in attachment to the substrate surface, etc. Desirable LNA units also are disclosed in WO 0056746, WO 0056748, and WO 30 0066604.

Desirable syntheses of pyrene-LNA monomers is shown in the following Schemes 1 and 2. In the below Schemes 1 and 2, the compound reference numerals are also referred to in the examples below.



Scheme 1



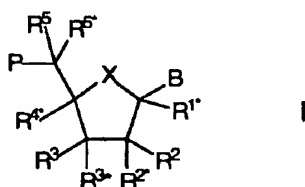
Scheme 2

A wide variety of modified nucleic acids may be employed, including those that have 2'-modification of hydroxyl, 2'-O-methyl, 2'-fluoro, 2'-trifluoromethyl, 2'-O-(2-methoxyethyl), 2'-O-aminopropyl, 2'-O-dimethylamino-oxyethyl, 2'-O-fluoroethyl or 2'-O-propenyl. The nucleic acid may further include a 3' modification, desirably where the 2'- and 3'-position of the 5 ribose group is linked. The nucleic acid also may contain a modification at the 4'-position, desirably where the 2'- and 4'-positions of the ribose group are linked such as by a 2'-4' link of -CH<sub>2</sub>-S-, -CH<sub>2</sub>-NH-, or -CH<sub>2</sub>-NMe- bridge.

The nucleotide also may have a variety of configurations such as  $\alpha$ -D-ribo,  $\beta$ -D-xylo, or  $\alpha$ -L-xylo configuration.

10 The internucleoside linkages of the units of oligos of the invention may be natural phosphodiester linkages, or other linkages such as -O-P(O)<sub>2</sub>-O-, -O-P(O,S)-O-, -O-P(S)<sub>2</sub>-O-, -NR<sup>H</sup>-P(O)<sub>2</sub>-O-, -O-P(O,NR<sup>H</sup>)-O-, -O-PO(R'')-O-, -O-PO(CH<sub>3</sub>)-O-, and -O-PO(NHR<sup>N</sup>)-O-, where R<sup>H</sup> is selected from hydrogen and C<sub>1-4</sub>-alkyl, and R'' is selected from C<sub>1-6</sub>-alkyl and phenyl.

15 A further desirable group of modified nucleic acids for incorporation into oligomers of the invention include those of the following formula:



20 wherein X is -O-; B is a modified base as discussed above e.g. an optionally substituted carbocyclic aryl such as optionally substituted pyrene or optionally substituted pyrenylmethylglycerol, or an optionally substituted heteroalicyclic or optionally substituted heteroaromatic such as optionally substituted pyridyloxazole. Other desirable universal bases include, pyrrole, diazole or triazole moieties, all of which may be optionally substituted. R<sup>1\*</sup> is 25 hydrogen.

P designates the radical position for an internucleoside linkage to a succeeding monomer, or a 5'-terminal group, such internucleoside linkage or 5'-terminal group optionally including the substituent R<sup>5</sup>, R<sup>5</sup> being hydrogen or included in an internucleoside linkage. R<sup>3\*</sup> is a group P\* which designates an internucleoside linkage to a preceding monomer, or a 3'-30 terminal group. One or two pairs of non-geminal substituents selected from the present substituents of R<sup>2</sup>, R<sup>2\*</sup>, R<sup>3</sup>, R<sup>4\*</sup>, may designate a biradical consisting of 1-4 groups/atoms

selected from  $-C(R^aR^b)-$ ,  $-C(R^a)=C(R^b)-$ ,  $-C(R^a)=N-$ ,  $-O-$ ,  $-S-$ ,  $-SO_2-$ ,  $-N(R^a)-$ , and  $>C=Z$ . Z is selected from  $-O-$ ,  $-S-$ , and  $-N(R^a)-$ , and  $R^a$  and  $R^b$  each is independently selected from hydrogen, optionally substituted  $C_{1-6}$ -alkyl, optionally substituted  $C_{2-6}$ -alkenyl, hydroxy,  $C_{1-6}$ -alkoxy,  $C_{2-6}$ -alkenyloxy, carboxy,  $C_{1-6}$ -alkoxycarbonyl,  $C_{1-6}$ -alkylcarbonyl, formyl, amino, mono- and di( $C_{1-6}$ -alkyl)amino, carbamoyl, mono- and di( $C_{1-6}$ -alkyl)-amino-carbonyl, amino- $C_{1-6}$ -alkyl-aminocarbonyl, mono- and di( $C_{1-6}$ -alkyl)amino- $C_{1-6}$ -alkyl-aminocarbonyl,  $C_{1-6}$ -alkyl-carbonylamino, carbamido,  $C_{1-6}$ -alkanoyloxy, sulphonyl,  $C_{1-6}$ -alkylsulphonyloxy, nitro, azido, sulphonyl,  $C_{1-6}$ -alkylthio, halogen, photochemically active groups, thermochemically active groups, chelating groups, reporter groups, and ligands, the possible pair of non-geminal substituents thereby forming a monocyclic entity together with (i) the atoms to which the non-geminal substituents are bound and (ii) any intervening atoms; and each of the substituents  $R^2$ ,  $R^{2*}$ ,  $R^3$ ,  $R^{4*}$  which are present and not involved in the possible biradical is independently selected from hydrogen, optionally substituted  $C_{1-6}$ -alkyl, optionally substituted  $C_{2-6}$ -alkenyl, hydroxy,  $C_{1-6}$ -alkoxy,  $C_{2-6}$ -alkenyloxy, carboxy,  $C_{1-6}$ -alkoxycarbonyl,  $C_{1-6}$ -alkylcarbonyl, formyl, amino, mono- and di( $C_{1-6}$ -alkyl)amino, carbamoyl, mono- and di( $C_{1-6}$ -alkyl)-amino-carbonyl, amino- $C_{1-6}$ -alkyl-aminocarbonyl, mono- and di( $C_{1-6}$ -alkyl)amino- $C_{1-6}$ -alkyl-aminocarbonyl,  $C_{1-6}$ -alkyl-carbonylamino, carbamido,  $C_{1-6}$ -alkanoyloxy, sulphonyl,  $C_{1-6}$ -alkylsulphonyloxy, nitro, azido, sulphonyl,  $C_{1-6}$ -alkylthio, halogen, photochemically active groups, thermochemically active groups, chelating groups, reporter groups, and ligands; and basic salts and acid addition salts thereof.

Modified nucleobases and nucleosidic bases may comprise a cyclic unit (e.g. a carbocyclic unit such as pyrenyl) that is joined to a nucleic unit, such as a 1'-position of furansyl ring through a linker, such as a straight or branched chain alkylene or alkenylene group. Alkylene groups suitably having from 1 (i.e.  $-CH_2-$ ) to about 12 carbon atoms, more typically 1 to about 8 carbon atoms, still more typically 1 to about 6 carbon atoms. Alkenylene groups suitably have one, two or three carbon-carbon double bonds and from 2 to 12 carbon atoms, more typically 2 to 8 carbon atoms, still more typically 2 to 6 carbon atoms.

#### Example 24: Exemplary Nucleic Acid Monomers and Oligomers

Desirable LNA units include those that contain a furanosyl-type ring and one or more of the following linkages: C-1', C-2'; C-2', C-3'; C-2', C-4'; or a C-2', C-5' linkage. A C-2', C-4' is particularly desirable. In another aspect of the invention, desirable LNA units are compounds having a substituent on the 2'-position of the central sugar moiety (e.g., ribose or xylose), or derivatives thereof, which favors the C3'-endo conformation, commonly referred to as the North (or simply N for short) conformation. Desirable LNA In various embodiments, the

oligonucleotide has at least one LNA unit with a modified base as disclosed herein. Suitable oligonucleotides also may contain natural DNA or RNA units (e.g., nucleotides) with natural bases, as well as LNA units that contain natural bases. Furthermore, the oligonucleotides of the invention also may contain modified DNA or RNA, such as 2'-O-methyl RNA, with natural or modified nucleobases (e.g., pyrene). Desirable oligonucleotides contain at least one of and desirably both of 1) one or more DNA or RNA units (e.g., nucleotides) with natural bases, and 2) one or more LNA units with natural bases, in addition to LNA units with a modified base. In other embodiments, the nucleic acid does not contain a modified base.

Oligonucleotides of the invention desirably contain at least 50 percent or more, more desirably 55, 60, 65, or 70 percent or more of non-modified or natural DNA or RNA units (e.g., nucleotides) or units other than LNA units based on the total number of units or residues of the oligo. A non-modified nucleic acid as referred to herein means that the nucleic acid upon incorporation into a 10-mer oligomer will not increase the  $T_m$  of the oligomer in excess of 1°C or 2°C. More desirably, the non-modified nucleic acid unit (e.g., nucleotide) is a substantially or completely "natural" nucleic acid, i.e. containing a non-modified base of uracil, cytosine, 5-methyl-cytosine, thymine, adenine or guanine and a non-modified pentose sugar unit of  $\beta$ -D-ribose (in the case of RNA) or  $\beta$ -D-2-deoxyribose (in the case of DNA).

Oligonucleotides of the invention suitably may contain only a single modified (i.e. LNA) nucleic acid unit, but desirably an oligonucleotide will contain 2, 3, 4 or 5 or more modified nucleic acid units. Typically desirable is where an oligonucleotide contains from about 5 to about 40 or 45 percent modified (LNA) nucleic acid units, based on total units of the oligo, more desirably where the oligonucleotide contains from about 5 or 10 percent to about 20, 25, 30 or 35 percent modified nucleic acid units, based on total units of the oligo.

Typical oligonucleotides that contain one or more LNA units with a modified base as disclosed herein suitably contain from 3 or 4 to about 200 nucleic acid repeat units, with at least one unit being an LNA unit with a modified base, more typically from about 3 or 4 to about 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140 or 150 nucleic acid units, with 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 LNA units with a modified base being present.

As discussed above, particularly desirable oligonucleotides contain a non-modified DNA or RNA unit at the 3' terminus and a modified DNA or RNA unit at one position upstream from (generally referred to hereing as the -1 or penultimate position) the 3' terminal non-modified nucleic acid unit. In some embodiments, the modified base is at the 3' terminal position of a nucleic acid primer, such as a primer for the detection of a single nucleotide polymorphism. Other particularly desirable nucleic acids have an LNA unit with or without a modified base in the 5' and/or 3' terminal position.

Also desirable are oligonucleotides that do not have an extended stretches of modified DNA or RNA units, e.g. greater than about 4, 5 or 6 consecutive modified DNA or RNA units. That is, desirably one or more non-modified DNA or RNA will be present after a consecutive stretch of about 3, 4 or 5 modified nucleic acids.

5 Generally desirable are oligonucleotides that contain a mixture of LNA units that have non-modified or natural nucleobases (i.e., adenine, guanine, cytosine, 5-methyl-cytosine, uracil, or thymine) and LNA units that have modified bases as disclosed herein.

Particularly desirable oligonucleotides of the invention include those where an LNA unit with a modified base is interposed between two LNA units each having non-modified or  
10 natural bases (adenine, guanine, cytosine, 5-methyl-cytosine, uracil, or thymine. The LNA "flanking" units with natural base moieties may be directly adjacent to the LNA with modified base moiety, or desirably is within 2, 3, 4 or 5 nucleic acid units of the LNA unit with modified base. Nucleic acid units that may be spaced between an LNA unit with a modified base and an LNA unit with natural nucleobasis suitably are DNA and/or RNA and/or alkyl-modified  
15 RNA/DNA units, typically with natural base moieties, although the DNA and or RNA units also may contain modified base moieties.

The oligonucleotides of the present invention are comprised of at least about one universal base. Oligonucleotides of the present can also be comprised, for exmple, of between about one to six 2'-Ome-RNA unit, at least about two LNA units and at least about one LNA  
20 pyrene unit.

#### Example 25: Exemplary Target Nucleic Acids

In the practice of the present invention, target genes may be suitably single-stranded or double-stranded DNA or RNA; however, single-stranded DNA or RNA targets are desirable. It  
25 is understood that the target to which the nucleic acids of the invention are directed include allelic forms of the targeted gene and the corresponding mRNAs including splice variants. There is substantial guidance in the literature for selecting particular sequences for nucleic acids with LNA or other high affinity nucleotides given a knowledge of the sequence of the target polynucleotide, e.g., Peyman and Ulmann, *Chemical Reviews*, 90:543-584, 1990; Crooke, *Ann.*  
30 *Rev. Pharmacol. Toxicol.*, 32:329-376 (1992); and Zamecnik and Stephenson, *Proc. Natl. Acad. Sci.*, 75:280-284 (1974). Desirable mRNA targets include the 5' cap site, tRNA primer binding site, the initiation codon site, the mRNA donor splice site, and the mRNA acceptor splice site, e.g., Goodchild et al., U.S. Patent 4,806,463.

#### 35 Example 26: Exemplary Applications of Present Methods



The chimeric oligos of the present invention are highly suitable for a variety of diagnostic purposes such as for the isolation, purification, amplification, detection, identification, quantification, or capture of nucleic acids such as DNA, mRNA or non-protein coding cellular RNAs, such as tRNA, rRNA, snRNA and scRNA, or synthetic nucleic acids, *in vivo* or *in vitro*.

The oligomer can comprise a photochemically active group, a thermochemically active group, a chelating group, a reporter group, or a ligand that facilitates the direct or indirect detection of the oligomer or the immobilization of the oligomer onto a solid support. Such group are typically attached to the oligo when it is intended as a probe for *in situ* hybridization, 10 in Southern hybridization, Dot blot hybridization, reverse Dot blot hybridization, or in Northern hybridization.

When the photochemically active group, the thermochemically active group, the chelating group, the reporter group, or the ligand includes a spacer (K), the spacer may suitably comprise a chemically cleavable group.

15 An additional object of the present invention is to provide oligonucleotides which combines an increased ability to discriminate between complementary and mismatched targets with the ability to act as substrates for nucleic acid active enzymes such as for example DNA and RNA polymerases, ligases, phosphatases. Such oligonucleotides may be used for instance as primers for sequencing nucleic acids and as primers in any of the several well known 20 amplification reactions, such as the PCR reaction.

Introduction of LNA monomers with natural bases into either DNA, RNA, or pure LNA oligonucleotides can result in extremely high thermal stability of duplexes with complimentary DNA or RNA, while at the same time obeying the Watson-Crick base-pairing rules. In general, the thermal stability of heteroduplexes is increased 3-8°C per LNA monomer in the duplex. 25 Oligonucleotides containing LNA can be designed to be substrates for polymerases (e.g., *Taq* polymerase), and PCR based on LNA primers is more discriminatory towards single base mutations in the template DNA compared to normal DNA-primers (e.g., allele specific PCR). Furthermore, very short LNA oligos (e.g. 5-mers or 8-mers) which have high  $T_m$ 's when compared to similar DNA oligos can be used as highly specific catching probes with 30 outstanding discriminatory power towards single base mutations (e.g., SNP detection).

LNA oligonucleotides are capable of hybridizing with double-stranded DNA target molecules as well as RNA secondary structures by strand invasion as well as of specifically blocking a wide selection of enzymatic reactions such as, digestion of double-stranded DNA by restriction endonucleases; and digestion of DNA and RNA with deoxyribonucleases and 35 ribonucleases, respectively.

In a further aspect, oligonucleotides of the invention may be used to construct new affinity pairs with exhibit enhanced specificity towards each other. The affinity constants can easily be adjusted over a wide range and a vast number of affinity pairs can be designed and synthesized. One part of the affinity pair can be attached to the molecule of interest (*e.g.* 5 proteins, amplicons, enzymes, polysaccharides, antibodies, haptens, peptides, etc.) by standard methods, while the other part of the affinity pair can be attached to *e.g.* a solid support such as beads, membranes, micro-titer plates, sticks, tubes, etc. The solid support may be chosen from a wide range of polymer materials such as for instance polypropylene, polystyrene, polycarbonate or polyethylene. The affinity pairs may be used in selective isolation, 10 purification, capture and detection of a diversity of the target molecules.

Oligonucleotides of the invention also may be employed as probes in the purification, isolation and detection of for instance pathogenic organisms such as viral, bacteria and fungi etc. Oligonucleotides of the invention also may be used as generic tools for the purification, isolation, amplification and detection of nucleic acids from groups of related species such as for 15 instance rRNA from gram-positive or gram negative bacteria, fungi, mammalian cells etc.

Oligonucleotides of the invention also may be employed as an aptamer in molecular diagnostics, *e.g.* in RNA mediated catalytic processes, in specific binding of antibiotics, drugs, amino acids, peptides, structural proteins, protein receptors, protein enzymes, saccharides, polysaccharides, biological cofactors, nucleic acids, or triphosphates or in the separation of 20 enantiomers from racemic mixtures by stereospecific binding.

Oligonucleotides of the invention also may be used for labeling of cells, *e.g.* in methods wherein the label allows the cells to be separated from unlabelled cells.

Oligonucleotides also may be conjugated to a compound selected from proteins, amplicons, enzymes, polysaccharides, antibodies, haptens, and peptides.

25 Kits are also provided containing one or more oligonucleotides of the invention for the isolation, purification, amplification, detection, identification, quantification, or capture of natural or synthetic nucleic acids. The kit typically will contain a reaction body, *e.g.* a slide or biochip. One or more oligonucleotides of the invention may be suitably immobilized on such a reaction body.

30 The invention also provides methods for using kits of the invention for carrying out a variety of bioassays, *e.g.* for diagnostic purposes. Any type of assay wherein one component is immobilized may be carried out using the substrate platforms of the invention. Bioassays utilizing an immobilized component are well known in the art. Examples of assays utilizing an immobilized component include for example, immunoassays, analysis of protein-protein 35 interactions, analysis of protein-nucleic acid interactions, analysis of nucleic acid-nucleic acid

interactions, receptor binding assays, enzyme assays, phosphorylation assays, diagnostic assays for determination of disease state, genetic profiling for drug compatibility analysis, and SNP detection (U.S.P.N. 6,316,198; 6,303,315).

Identification of a nucleic acid sequence capable of binding to a biomolecule of interest 5 can be achieved by immobilizing a library of nucleic acids onto the substrate surface so that each unique nucleic acid was located at a defined position to form an array. The array would then be exposed to the biomolecule under conditions which favored binding of the biomolecule to the nucleic acids. Non-specifically binding biomolecules could be washed away using mild to stringent buffer conditions depending on the level of specificity of binding desired. The 10 nucleic acid array would then be analyzed to determine which nucleic acid sequences bound to the biomolecule. Desirably the biomolecules would carry a fluorescent tag for use in detection of the location of the bound nucleic acids.

Oligonucleotides of the invention can be employed in a wide range of applications, particularly those in those applications involving a hybridization reaction. Oligonucleotides 15 also may be used in DNA sequencing aiming at improved throughput in large-scale, shotgun genome sequencing projects, improved throughput in capillary DNA sequencing (e.g. ABI prism 3700) as well as at an improved method for 1) sequencing large, tandemly repeated genomic regions, 2) closing gaps in genome sequencing projects and 3) sequencing of GC-rich templates. In DNA sequencing, oligonucleotide sequencing primers are combined with LNA 20 enhancer elements for the read-through of GC-rich and/or tandemly repeated genomic regions, which often present many challenges for genome sequencing projects. LNA may increase the specificity of certain sequencing primers and thus facilitate selection of a particular version of a repeated sequence and possibly also use strand invasion to open up recalcitrant GC rich sequences.

25 The incorporation of one or more universal nucleosides into the oligomer makes bonding to unknown bases possible and allows the oligonucleotide to match ambiguous or unknown nucleic acid sequences.

As discussed above, oligonucleotides of the invention may be used for therapeutic applications, e.g. as an antisense, antigene or ribozyme or double stranded nucleic acid 30 therapeutic agents. In these therapeutic methods, one or more oligonucleotides of the invention is administered as desired to a patient suffering from or susceptible the targeted disease or disorder, e.g. a viral infection.

In an exemplary *in vitro* method for measuring the ability of a nucleic acid of the invention to silence a target gene, cells are cultured in standard medium supplemented with 1% 35 fetal calf serum as previously described (Lykkesfeld *et al.*, Int. J. Cancer 61:529-534, 1995). At

the start of the experiment cells are approximately 40% confluent. The serum containing medium is removed and replaced with serum-free medium. Transfection is performed using, e.g., Lipofectin (GibcoBRL cat. No 18292-011) diluted 40X in medium without serum and combined with the oligo to a concentration of 750 nM oligo, 0.8 ug/ml Lipofectin. Then, the medium is removed from the cells and replaced with the medium containing oligo-Lipofectin complex. The cells are incubated at 37°C for 6 hours, rinsed once with medium without serum and incubated for a further 18 hours in DME/F12 with 1% FCS at 37°C. Standard methods are used to measure the level of mRNA or protein encoded by the target gene to measure the level of gene silencing.

10 It is also contemplated that information on the structures assumed by a target nucleic acid may be used in the design of the probes, such that regions that are known or suspected to be involved in folding may be chosen as hybridization sites. Such an approach will reduce the number of probes that are likely to be needed to distinguish between targets of interest.

There are many methods used to obtain structural information involving nucleic acids, 15 including the use of chemicals that are sensitive to the nucleic acid structure, such as phenanthroline/copper, EDTA-Fe<sup>2+</sup>, cisplatin, ethylnitrosourea, dimethyl pyrocarbonate, hydrazine, dimethyl sulfate, and bisulfite. Enzymatic probing using structure-specific nucleases from a variety of sources, such as the Cleavase™ enzymes (Third Wave Technologies, Inc., Madison, Wis.), Taq DNA polymerase, E. coli DNA polymerase I, and eukaryotic structure- 20 specific endonucleases (e.g., human, murine and Xenopus XPG enzymes, yeast RAD2 enzymes), murine FEN-1 endonucleases (Harrington and Lieber, Genes and Develop., 3:1344 [1994]) and calf thymus 5' to 3' exonuclease (Murante et al., J. Biol. Chem., 269:1191 [1994]). In addition, enzymes having 3' nuclease activity such as members of the family of DNA repair endonucleases (e.g., the RrpI enzyme from Drosophila melanogaster, the yeast RAD1/RAD10 25 complex and E. coli Exo III), are also suitable for examining the structures of nucleic acids.

If analysis of structure as a step in probe selection is to be used for a segment of nucleic acid for which no information is available concerning regions likely to form secondary structures, the sites of structure-induced modification or cleavage must be identified. It is most convenient if the modification or cleavage can be done under partially reactive conditions (i.e., 30 such that in the population of molecules in a test sample, each individual will receive only one or a few cuts or modifications). When the sample is analyzed as a whole, each reactive site should be represented, and all the sites may be thus identified. Using a Cleavase Fragment Length Polymorphism™ cleavage reaction as an example, when the partial cleavage products of an end labeled nucleic acid fragment are resolved by size (e.g., by electrophoresis), the result 35 is a ladder of bands indicating the site of each cleavage, measured from the labeled end. Similar

analysis can be done for chemical modifications that block DNA synthesis; extension of a primer on molecules that have been partially modified will yield a nested set of termination products. Determining the sites of cleavage/modification may be done with some degree of accuracy by comparing the products to size markers (e.g., commercially available fragments of 5 DNA for size comparison) but a more accurate measure is to create a DNA sequencing ladder for the same segment of nucleic acid to resolve alongside the test sample. This allows rapid identification of the precise site of cleavage or modification.

Example 27: General Reaction Conditions for Synthesis of Some Compounds of the Invention

10 Reactions were conducted under an atmosphere of nitrogen when anhydrous solvents were used. All reactions were monitored by thin-layer chromatography (TLC) using EM reagent plates with fluorescence indicator (SiO<sub>2</sub>-60, F-254). The compounds were visualized under UV light and by spraying with a mixture of 5% aqueous sulfuric acid and ethanol followed by heating. Silica gel 60 (particle size 0.040-0.063 mm, Merck) was used for flash  
15 column chromatography. NMR spectra were recorded at 300 MHz for <sup>1</sup>H NMR, 75.5 MHz for <sup>13</sup>C NMR and 121.5 MHz for <sup>31</sup>P NMR on a Varian Unity 300 spectrometer.  $\delta$ -Values are in ppm relative to tetramethyl silane as internal standard (<sup>1</sup>H and <sup>13</sup>C NMR) and relative to 85% H<sub>3</sub>PO<sub>4</sub> as external standard (<sup>31</sup>P NMR). Coupling constants are given in Hertz. The assignments, when given, are tentative, and the assignments of methylene protons, when given,  
20 may be interchanged. Bicyclic compounds are named according to the Von Bayer nomenclature. Fast atom bombardment mass spectra (FAB-MS) were recorded in positive ion mode on a Kratos MS50TC spectrometer. The composition of the oligonucleotides were verified by MALDI-MS on a Micromass Tof Spec E mass spectrometer using a matrix of diammonium citrate and 2,6-dihydroxyacetophenone.

25

Example 28: Synthesis of 1,2-O-Isopropylidene-5-O-methanesulfonyl-4-C-methanesulfonyloxymethyl-3-O-(p-methoxybenzyl)- $\alpha$ -D-ribofuranose [Compound 2 in Scheme 1 above]

Mesyl chloride (8.6 g, 7.5 mmol) was dropwise added to a stirred solution of 4-C-  
30 hydroxymethyl-1,2-O-isopropylidene-3-O-p-methoxybenzyl- $\alpha$ -D-ribofuranose [R. Yamaguchi, T. Imanishi, S. Kohgo, H. Horie and H. Ohnui, *Biosci. Biotechnol. Biochem.*, 1999, **63**, 736] (1, 10.0 g, 29.4 mmol) in anhydrous pyridine (30 cm<sup>3</sup>) and the reaction mixture was stirred overnight at room temperature. The mixture was evaporated to dryness under reduced pressure to give a residue which was co-evaporated with toluene (2 x 25 cm<sup>3</sup>), dissolved in CH<sub>2</sub>Cl<sub>2</sub> (200

cm<sup>3</sup>) and washed successively with saturated aqueous NaHCO<sub>3</sub> (2 x 100 cm<sup>3</sup>) and brine (50 cm<sup>3</sup>). The organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>), filtered and evaporated to dryness under reduced pressure. The colorless viscous oil obtained was purified by column chromatography [0.5-1% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub> as eluent], followed by crystallization from MeOH to give furanose 2 as 5 a white solid material (13.6 g, 93%); *R<sub>f</sub>* 0.57 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 7.30 (2 H, d, *J* 8.7), 6.90 (2 H, d, *J* 8.5), 5.78 (1 H, d, *J* 3.7), 4.86 (1 H, d, *J* 12.0), 4.70 (1 H, d, *J* 11.4), 4.62 (1 H, dd, *J* 5.0 and 3.8), 4.50 (1 H, d, *J* 11.1), 4.39 (1 H, d, *J* 12.3), 4.31 (1 H, d, *J* 11.0), 4.17 (1 H, d, *J* 5.1), 4.11 (1 H, d, *J* 11.0), 3.81 (3 H, s), 3.07 (3 H, s), 2.99 (3 H, s), 1.68 (3 H, s), 1.34 (3 H, s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 159.8, 129.9, 128.8, 114.1, 114.0, 104.5, 83.2, 78.0, 77.9, 72.6, 10 69.6, 68.8, 55.4, 38.1, 37.5, 26.3, 25.7.

Example 29: Synthesis of Methyl 5-O-methanesulfonyl-4-C-methanesulfonyloxymethyl-3-O-(p-methoxybenzyl)-D-ribofuranoside [Compound 3 in Scheme 1 above]

A suspension of furanoside 2 (13.5 g, 27.2 mmol) in a mixture of H<sub>2</sub>O (45 cm<sup>3</sup>) and 15 15% HCl in MeOH (450 cm<sup>3</sup>, w/w) was stirred at room temperature for 72 h. The mixture was carefully neutralized by addition of saturated aqueous NaHCO<sub>3</sub> (100 cm<sup>3</sup>) followed by NaHCO<sub>3</sub> (s) whereupon the mixture was evaporated to dryness under reduced pressure. H<sub>2</sub>O (100cm<sup>3</sup>) was added, and extraction was performed with EtOAc (3 x 100 cm<sup>3</sup>). The combined organic phase was washed with brine (100 cm<sup>3</sup>), dried (Na<sub>2</sub>SO<sub>4</sub>), filtered and then evaporated to dryness 20 under reduced pressure. The residue was coevaporated with toluene (2 x 25 cm<sup>3</sup>) and purified by column chromatography [1-2% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give furanoside 3 as an anomeric mixture (clear oil, 11.0 g, 86%, ratio between anomers ca. 6:1); *R<sub>f</sub>* 0.39, 0.33 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>, major anomer only) 7.28 (2 H, d, *J* 8.4), 6.91 (2 H, d, *J* 8.9), 4.87 (1 H, s), 4.62 (1 H, d, *J* 11.4), 4.53 (1 H, d, *J* 11.2), 4.41 (2 H, s), 4.31 (1 H, d, *J* 9.8), 4.24 (1 H, d, *J* 25 4.6), 4.06 (1 H, d, *J* 10.0), 3.98 (1 H, br s), 3.81 (3 H, s), 3.33 (3 H, s), 3.06 (3 H, s), 3.03 (3 H,s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>, major anomer only) 160.0, 130.1, 128.5, 114.3, 107.8, 81.7, 81.2, 73.8, 73.6, 69.7, 69.6, 55.5, 55.4, 37.5, 37.4.

Example 30: Synthesis of (1R,3RS,4R,7S)-1-Methanesulfonyloxymethyl-3-methoxy-7-(p-methoxybenzyloxy)-2,5-dioxabicyclo[2.2.1]heptane [Compound 4 in Scheme 1 above]

To a stirred solution of the anomeric mixture of Compound 3 (10.9 g, 23.2 mmol) in anhydrous DMF (50 cm<sup>3</sup>) at 0 °C was during 10 min added sodium hydride (2.28 g of a 60% suspension in mineral oil (w/w), 95.2 mmol) and the mixture was stirred for 12 h at room temperature. Ice-cold H<sub>2</sub>O (200 cm<sup>3</sup>) was slowly added and extraction was performed using

EtOAc (3 x 200 cm<sup>3</sup>). The combined organic phase was washed successively with saturated aqueous NaHCO<sub>3</sub> (2 x 100 cm<sup>3</sup>) and brine (50 cm<sup>3</sup>), dried (Na<sub>2</sub>SO<sub>4</sub>), filtered and evaporated to dryness under reduced pressure. The residue was purified by column chromatography [0.5-1% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub> ] to give first the major isomer (6.42 g, 74%) and then [1.5% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub> ] the minor isomer (1.13 g, 13%), both as clear oils; *R<sub>f</sub>* 0.56, 0.45 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>, major isomer) 7.16 (2 H, d, *J* 8.8), 6.74 (2 H, d, *J* 8.4), 4.65 (1 H, s), 4.42-4.32 (4 H, m), 3.95-3.94 (2 H, m), 3.84 (1 H, d, *J* 7.4), 3.66 (3 H, s), 3.54 (1 H, d, *J* 7.4), 3.21 (3 H, s), 2.90 (3 H, s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>, major isomer) 159.6, 129.5, 129.3, 114.0, 105.3, 83.2, 78.6, 77.2, 72.1, 71.8, 66.3, 55.6, 55.4, 37.8;  $\delta_{\text{H}}$  (CDCl<sub>3</sub>, minor isomer) 7.27 (2 H, d, *J* 8.9), 6.89 (2 H, d, *J* 8.6), 4.99 (1 H, s), 4.63-4.39 (4 H, m), 4.19 (1 H, s), 4.10-3.94 (2 H, m), 3.91 (1 H, s), 3.81 (3 H, s), 3.47 (3 H, s), 3.05 (3 H, s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>, minor isomer) 159.7, 129.6, 129.5, 114.1, 104.4, 86.4, 79.3, 77.1, 72.3, 71.9, 66.2, 56.4, 55.4, 37.7.

**Example 31: Synthesis of (1*R*,4*R*,7*S*)-1-Acetoxymethyl-3-methoxy-7-(*p*-methoxybenzyloxy)-15 2,5-dioxabicyclo[2.2.1]heptane [Compound 5 in Scheme 1]**

To a stirred solution of furanoside 4 (major isomer, 6.36 g, 17.0 mmol) in dioxane (25 cm<sup>3</sup>) was added 18-crown-6 (9.0 g, 34.1 mmol) and KOAc (8.4 g, 85.6 mmol). The stirred mixture was heated under reflux for 12 h and subsequently evaporated to dryness under reduced pressure. The residue was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (100 cm<sup>3</sup>) and washing was performed, 20 successively, with saturated aqueous NaHCO<sub>3</sub> (2 x 50 cm<sup>3</sup>) and brine (50 cm<sup>3</sup>). The separated organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>), filtered and evaporated to dryness under reduced pressure. The residue was purified by column chromatography [1% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give furanoside 5 as a white solid material (one anomer, 5.23 g, 91%); *R<sub>f</sub>* 0.63 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 7.27-7.24 (2 H, m), 6.90-6.87 (2 H, m), 4.79 (1 H, s), 4.61 (1 H, d, *J* 11.0), 25 4.49 (2 H, m), 4.28 (1 H, d, *J* 11.0), 4.04 (3 H, m), 3.80 (3 H, s), 3.68 (1 H, m), 3.36 (3 H, s), 2.06 (3 H, s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 170.7, 159.5, 129.5, 129.4, 113.9, 105.1, 83.3, 78.9, 77.2, 72.0, 71.9, 61.0, 55.4, 55.3, 20.8.

**Example 32: Synthesis of (1*S*,4*R*,7*S*)-1-Hydroxymethyl-3-methoxy-7-(*p*-methoxybenzyloxy)-30 2,5-dioxabicyclo[2.2.1]heptane [Compound 6 in Scheme 1]**

A solution of furanoside 5 (one anomer, 5.16 g, 15.3 mmol) in saturated methanolic ammonia (200 cm<sup>3</sup>) was stirred at room temperature for 48 h. The reaction mixture was evaporated to dryness under reduced pressure, coevaporated with toluene (2 x 50 cm<sup>3</sup>), and the residue purified by column chromatography [2-3% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub> ] to give furanoside

**6** as a white solid material (one anomer, 3.98 g, 88%);  $R_f$  0.43 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  95:5, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 7.27 (2 H, d,  $J$  8.6), 6.88 (2 H, d,  $J$  8.9), 4.79 (1 H, s), 4.59 (1 H, d,  $J$  11.3), 4.53 (1 H, d,  $J$  11.4), 4.09 (2 H, s), 3.97 (1 H, d,  $J$  7.5), 3.86 (2 H, br s), 3.80 (3 H, s), 3.75-3.62 (2 H, m), 3.37 (3 H, s);  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 159.4, 129.7, 129.3, 113.9, 105.2, 85.6, 78.3, 77.4, 71.9, 71.8, 58.8, 55.5, 55.3.

Example 33: (1S,4R,7S)-3-Methoxy-7-(p-methoxybenzyloxy)-1-(p-methoxybenzyloxymethyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 7 in Scheme 1]

To a stirred solution of furanoside **6** (one anomer, 3.94 g, 13.3 mmol) in anhydrous DMF (50  $\text{cm}^3$ ) at 0 °C was added a suspension of NaH [60% in mineral oil (w/w), 1.46 g, 60.8 mmol] followed by dropwise addition of *p*-methoxybenzyl chloride (2.74 g, 17.5 mmol). The mixture was allowed to warm to room temperature and stirring was continued for another 4 h whereupon ice-cold  $\text{H}_2\text{O}$  (50  $\text{cm}^3$ ) was dropwise added. The mixture was extracted with  $\text{CH}_2\text{Cl}_2$  (3 x 100  $\text{cm}^3$ ) and the combined organic phase was washed with brine (100  $\text{cm}^3$ ), dried ( $\text{Na}_2\text{SO}_4$ ), filtered, evaporated to dryness under reduced pressure and coevaporated with toluene (3 x 50  $\text{cm}^3$ ). The residue (4.71 g) tentatively assigned as a mixture of **7** and aldehyde **11** was used in the preparation of **11** (see below) without further purification.

Example 34: 4-C-Methanesulfonyloxymethyl-3,5-di-O-(p-methoxybenzyl)-1,2-O-isopropylidene- $\alpha$ -D-ribofuranose [Compound 9 in Scheme 1]

4-C-Hydroxymethyl-3,5-di-O-(*p*-methoxybenzyl)-1,2-O-isopropylidene- $\alpha$ -D-ribofuranose [R. Yamaguchi, T. Imanishi, S. Kohgo, H. Horie and H. Ohrai, *Biosci. Biotechnol. Biochem.*, 1999, **63**, 736] (**8**, 3.2 g, 6.95 mmol) was mesylated using  $\text{MsCl}$  (2.00 g, 17.5 mmol) and pyridine (10  $\text{cm}^3$ ) following the procedure described for **2**. After work-up, the colorless viscous oil was purified by column chromatography [1% (v/v)  $\text{MeOH}$  in  $\text{CH}_2\text{Cl}_2$ ] to give derivative **9** in 89% yield (3.17 g) as a clear oil;  $R_f$  0.45 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  98:2, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 7.22 (2 H, d,  $J$  8.9), 7.18 (2 H, d,  $J$  8.7), 6.86 (4 H, d,  $J$  8.3), 5.76 (1 H, d,  $J$  3.8), 4.83 (1 H, d,  $J$  12.0), 4.64 (1 H, d,  $J$  11.6), 4.59 (1 H, m), 4.49-4.35 (4 H, m), 4.24 (1 H, d,  $J$  5.3), 3.80 (6 H, s), 3.56 (1 H, d,  $J$  10.5), 3.45 (1 H, d,  $J$  10.5), 3.06 (3 H, s), 1.67 (3 H, s), 1.33 (3 H, s);  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 159.6, 159.4, 129.9, 129.8, 129.7, 129.5, 129.4, 129.3, 114.0, 113.9, 113.8, 113.7, 113.6, 104.5, 84.9, 78.6, 78.1, 73.4, 72.4, 71.0, 69.9, 55.3, 38.0, 26.4, 25.9.

Example 35: Methyl 4-C-methanesulfonyloxymethyl-3,5-di-O-(p-methoxybenzyl)-D-ribofuranose [Compound 10 in Scheme 1]



Methanolysis of furanoside **9** (3.1 g, 5.76 mmol) was performed using a mixture of a solution of 15% HCl in MeOH (w/w, 120 cm<sup>3</sup>) and H<sub>2</sub>O (12 cm<sup>3</sup>) following the procedure described for the synthesis of **3**. After work-up, the crude product was purified by column chromatography [0.5-1% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give the major anomer of **10** (1.71 g, 58%) and [1-1.5% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] the minor anomer of **10** (0.47 g, 16%), both as clear oils; *R<sub>f</sub>* 0.31, 0.24 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v);  $\delta$  (major anomer, CDCl<sub>3</sub>) 159.8, 159.5, 129.9, 129.8, 129.6, 129.5, 129.0, 114.2, 114.1, 114.0, 113.9, 107.9, 84.7, 79.9, 74.2, 73.5, 73.5, 70.2, 64.4, 55.6, 55.4, 37.4.

10 Example 36: Alternative preparation of Compound **7** in Scheme 1

Ring closure of furanoside **10** (major anomer, 1.68 g, 3.28 mmol) was achieved using NaH (60% suspension in mineral oil (w/w), 0.32 g, 13.1 mmol) in anhydrous DMF (10 cm<sup>3</sup>) following the procedure described for the synthesis of **4** to give a crude product tentatively assigned as a mixture of furanoside **7** and aldehyde **11** (see below) (1.13 g).

15

Example 37: (2R,3S,4S)-4-Hydroxy-3-(p-methoxybenzyloxy)-4-(p-methoxybenzyloxymethyl)-tetrahydrofuran-2-carbaldehyde [Compound **11** in Scheme 1]

A solution of crude furanoside **7** (as a mixture with **11** as prepared as described above, 5.80 g) in 80% glacial acetic acid (100 cm<sup>3</sup>) was stirred at 50 °C for 4 h. The solvent was 20 distilled off under reduced pressure and the residue was successively coevaporated with absolute ethanol (3 x 25 cm<sup>3</sup>) and toluene (2 x 25 cm<sup>3</sup>) and purified by column chromatography [4-5% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give aldehyde **11** as a colorless oil (4.60 g); *R<sub>f</sub>* 0.37 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta$ <sub>H</sub> (CDCl<sub>3</sub>) 9.64 (1 H, br s), 7.27-7.17 (4 H, m), 6.87-6.84 (4 H, m), 4.59 (1 H, d, *J* 11.6), 4.51-4.41 (2 H, m), 4.35 (1 H, s), 3.92-3.90 (2 H, m), 3.79 (6 H, s), 3.77-25 3.68 (3 H, m), 3.55 (2 H, br s);  $\delta$  (CDCl<sub>3</sub>) 203.6, 159.5, 159.4, 129.7, 129.6, 129.5, 129.2, 114.0, 113.9, 113.8, 87.3, 86.7, 81.0, 75.1, 73.4, 71.6, 67.6, 55.3.

Example 38: General procedure for the reaction of aryl magnesium bromides with aldehyde **11** to give Compounds **12a-e** in Scheme 2

30 A solution of aldehyde **11** (Scheme 2) in anhydrous THF (10 cm<sup>3</sup>) was added dropwise during 5 min to a stirred solution of the aryl magnesium bromide dissolved in anhydrous THF at 0 °C. The mixture was allowed to heat to room temperature and stirred for 12 h. The mixture was evaporated to dryness under reduced pressure and the residue diluted with CH<sub>2</sub>Cl<sub>2</sub> and washed several times with saturated aqueous NH<sub>4</sub>Cl. The organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>),

filtered, and evaporated to dryness under reduced pressure. Column chromatography of the crude product obtained afforded the compounds **12a-e** as shown in Scheme 2.

Example 38a: Synthesis of (2S,3S,4S)-4-Hydroxy-2-[(R)-hydroxy(phenyl)methyl]-4-(p-methoxybenzyloxy)-3-(p-methoxybenzyloxymethyl) tetrahydrofuran [Compound 12a of Scheme 2]

Grignard reaction of phenylmagnesium bromide (1.0 M solution in THF, 14.2 cm<sup>3</sup>, 14.2 mmol) with aldehyde **11** (515 mg, 1.28 mmol) afforded **12a** as shown in Scheme 2. The crude product was purified by column chromatography [4% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give 10 tetrahydrofuran **12a** (540 mg, 88%) as a colorless oil; *R<sub>f</sub>* 0.34 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 7.40-7.19 (7 H, m), 6.91-6.73 (6 H, m), 4.73 (1 H, d, *J* 6.4), 4.48 (2 H, s), 4.08 (2 H, s), 3.88 (1 H, d, *J* 9.4), 3.79 (1 H, m), 3.78 (3 H, s), 3.76 (3 H, s), 3.75-3.69 (2 H, m), 3.50 (1 H, d, *J* 9.4), 3.45 (1 H, s), 3.42 (1 H, br s), 3.26 (1 H, br s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 159.5, 159.3, 140.7, 129.7, 129.6, 129.5, 129.2, 128.5, 128.0, 127.3, 113.9, 113.8, 113.7, 89.4, 84.6, 81.8, 75.3, 74.7, 73.5, 15 71.6, 69.3, 55.3; *m/z* (FAB) 503 [M+Na]<sup>+</sup>, 479 [M-H]<sup>+</sup>, 461 [M-H-H<sub>2</sub>O]<sup>+</sup>.

Example 38b: Synthesis of (2S,3S,4S)-4-Hydroxy-2-[(R)-hydroxy(4-fluoro-3-methylphenyl)methyl]-4-(p-methoxybenzyloxy)-3-(p-methoxybenzyloxymethyl)tetrahydrofuran [Compound 12b of Scheme 2]

20 Grignard reaction of 4-fluoro-3-methylphenylmagnesium bromide (1.0 M solution in THF, 15.0 cm<sup>3</sup>, 15.0 mmol) with aldehyde **11** (603 mg, 1.5 mmol) afforded **12b** as shown in Scheme 2. The crude product was purified by column chromatography [4-5% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give tetrahydrofuran **12b** (611 mg, 85%) as a colorless oil; *R<sub>f</sub>* 0.34 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 7.24-7.12 (5 H, m), 6.98-6.84 (5 H, m), 6.77 (1 H, d, *J* 8.5), 4.65 (1 H, 25 dd, *J* 2.8 and 6.4), 4.49 (2 H, s), 4.15 (2 H, s), 4.01 (1 H, dd, *J* 2.3 and 6.5), 3.87 (1 H, d, *J* 9.3), 3.79 (3H, s), 3.78 (3 H, s), 3.76-3.68 (2 H, m), 3.52 (1 H, s), 3.47 (1 H, d, *J* 10.3), 3.42 (1 H, d, *J* 2.9), 3.22 (1 H, s), 2.24 (3 H, d, *J* 0.8);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 162.7, 159.5, 159.4, 136.2, 136.1, 130.3, 130.2, 129.7, 129.6, 129.5, 129.4, 129.1, 126.1, 126.0, 115.1, 114.8, 114.0, 113.9, 113.8, 89.3, 84.5, 81.8, 75.3, 74.0, 73.5, 71.7, 69.2, 55.4, 55.3, 14.7 (d, *J* 3.9); *m/z* (FAB) 535 [M+Na]<sup>+</sup>, 511 30 [M-H]<sup>+</sup>, 493 [M-H-H<sub>2</sub>O]<sup>+</sup>.

Example 38c: Synthesis of (2S,3S,4S)-4-Hydroxy-2-[(R)-hydroxy(1-naphtyl)methyl]-4-(p-methoxybenzyloxy)-3-(p-methoxybenzyloxymethyl) tetrahydrofuran [Compound 12c of Scheme 2]

1-Bromonaphthalene (1.55 g, 7.5 mmol) was added to a stirred mixture of magnesium turnings (182 mg, 7.5 mmol) and iodine (10 mg) in THF (10 cm<sup>3</sup>). The mixture was stirred at 40 °C for 1 h whereupon it was allowed to cool to room temperature. A solution of aldehyde **11** (603 mg, 1.5 mmol) in THF (10 cm<sup>3</sup>) was added slowly and the reaction was stirred for 12 h. The crude product was purified by column chromatography [4-5% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give tetrahydrofuran **12c** (756 mg, 95%) as a colorless oil; *R*<sub>f</sub> 0.35 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 8.08 (1 H, m), 7.86 (1 H, m), 7.79 (1 H, d, *J* 8.2), 7.72 (1 H, d, *J* 7.2), 7.49-7.44 (3H, m), 7.18 (2 H, d, *J* 8.4), 6.84 (2 H, d, *J* 8.6), 6.74 (2 H, d, *J* 8.7), 6.68 (2 H, d, *J* 8.8), 5.52 (1 H, dd, *J* 3.7 and 5.6), 4.45 (2 H, s), 4.34 (1 H, dd, *J* 2.5 and 5.9), 4.03 (1 H, d, *J* 11.0), 3.96 (1 H, d, *J* 11.0), 3.93 (1 H, d, *J* 9.5), 3.80 (1 H, d, *J* 9.3), 3.77 (3 H, s), 3.75 (1 H, d, *J* 2.6), 3.72 (3 H, s), 3.68 (1 H, d, *J* 9.3), 3.56 (1 H, d, *J* 3.7), 3.49 (1 H, d, *J* 9.3), 3.34 (1 H, s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 159.5, 159.3, 136.3, 134.0, 131.0, 129.7, 129.6, 129.5, 129.4, 129.0, 128.6, 128.2, 125.6, 125.5, 123.5, 114.0, 113.8, 113.7, 88.7, 84.7, 81.9, 75.5, 73.5, 71.7, 71.3, 69.3, 55.4, 55.3; *m/z* (FAB) 553 [M+Na]<sup>+</sup>, 529 [M-H]<sup>+</sup>, 511 [M-H-H<sub>2</sub>O]<sup>+</sup>.

15

Example 38d: (2S,3S,4S)-4-Hydroxy-2-[(R)-hydroxy(1-pyrenyl)methyl]-4-(p-methoxybenzyloxy)-3-(p-methoxybenzyloxymethyl)tetrahydrofuran [Compound 12d of Scheme 2]

Tetrahydrofuran **12d** was synthesized from aldehyde **11** (515 mg, 1.28 mmol), 1-bromopyrene (1.0 g, 3.56 mmol), magnesium turnings (155 mg, 6.4 mmol), iodine (10 mg) and THF (20 cm<sup>3</sup>) following the procedure described for synthesis of compound **12c**. The crude product was purified by column chromatography [3-4% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give tetrahydrofuran **12d** (690 mg, 89%) as a pale yellow solid; *R*<sub>f</sub> 0.35 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 8.23 (2 H, d, *J* 8.4 and 9.2), 8.19-8.13 (3 H, m), 8.05-7.99 (4 H, m), 7.14 (2 H, d, *J* 8.8), 6.82 (2 H, d, *J* 9.0), 6.30 (2 H, d, *J* 8.7), 6.20 (2 H, d, *J* 8.6), 5.87 (1 H, d, *J* 7.2), 4.43 (2 H, s), 4.41 (1 H, m), 4.01 (1 H, d, *J* 9.4), 3.91 (1 H, d, *J* 11.8), 3.86 (1 H, d, *J* 9.2), 3.77 (1 H, d, *J* 1.9), 3.76 (3 H, s), 3.70-3.64 (3 H, m), 3.52-3.45 (1 H, m), 3.44 (3 H, s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 159.5, 158.9, 133.9, 131.4, 131.1, 130.7, 129.7, 129.5, 129.2, 128.9, 128.5, 127.8, 127.7, 127.5, 126.0, 125.5, 125.3, 125.2, 125.1, 125.0, 124.9, 122.9, 113.9, 113.3, 89.5, 83.5, 82.0, 75.7, 73.4, 71.3, 71.0, 69.3, 55.3, 55.0; *m/z* (MALDI) 627 [M+Na]<sup>+</sup>, 609 [M<sup>+</sup>+Na-H<sub>2</sub>O]<sup>+</sup>.

30

Example 38e: (2S,3S,4S)-4-Hydroxy-2-[(R)-hydroxy(2,4,5-trimethylphenyl)methyl]-4-(p-methoxybenzyloxy)-3-(p-methoxybenzyloxymethyl) tetrahydrofuran [Compound 12e of Scheme 2]

Tetrahydrofuran **12e** was synthesized from aldehyde **11** (515 mg, 1.28 mmol), 1-bromo-2,4,5-trimethylbenzene (1.28 g, 6.4 mmol), magnesium turnings (155 mg, 6.4 mmol), iodine (10 mg) and THF (20 cm<sup>3</sup>) following the procedure described for synthesis of compound **12c**.

The crude product was purified by column chromatography [3-4% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give tetrahydrofuran **12e** (589 mg, 88%) as a colorless oil; *R*<sub>f</sub> 0.34 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 95:5, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 7.25 (2 H, d, *J* 8.7), 7.21 (2 H, d, *J* 8.9), 6.90 (1 H, s), 6.87 (1 H, s), 6.85 (2 H, d, *J* 8.9), 6.76 (2 H, d, *J* 8.7), 4.95 (1 H, dd, *J* 3.6 and 5.9), 4.48 (2 H, s), 4.18-4.08 (3 H, m), 3.89 (1 H, d, *J* 9.6), 3.80 (1 H, m), 3.79 (3 H, s), 3.77 (3 H, s), 3.71 (1 H, d, *J* 9.2), 3.64 (1 H, d, *J* 2.6), 3.51 (1 H, d, *J* 9.4), 3.24 (1 H, s), 3.18 (1 H, d, *J* 3.4), 2.25 (3 H, s), 2.22 (3 H, s), 2.21 (3 H, s);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 159.5, 159.3, 136.0, 135.8, 134.2, 132.5, 132.0, 129.8, 129.7, 129.6, 129.5, 128.5, 113.9, 113.8, 88.6, 84.7, 81.7, 75.4, 73.5, 71.7, 70.9, 69.4, 55.3, 19.5, 19.4, 19.0; *m/z* (FAB) 545 [M+Na]<sup>+</sup>, 521 [M-H]<sup>+</sup>, 503 [M-H-H<sub>2</sub>O]<sup>+</sup>.

Example 39: General procedure for the cyclization of **12a-e** to give compounds **13a-e** as shown in Scheme 2.

*N,N,N',N'*-Tetramethylazodicarboxamide (TMAD) was added in one portion to a stirred solution of the compounds **12a-e** as shown in Scheme 2 and tributylphosphine in benzene at 0 °C. The mixture was stirred for 12 h at room temperature whereupon it was diluted with diethyl ether (50 cm<sup>3</sup>). The organic phase was washed successively with saturated aqueous NH<sub>4</sub>Cl (2 x 20 cm<sup>3</sup>) and brine (25 cm<sup>3</sup>), dried (Na<sub>2</sub>SO<sub>4</sub>), filtered and evaporated to dryness under reduced pressure. The crude product obtained was purified by column chromatography [1.5-2% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>] to give compounds **13a-e** as shown in Scheme 2.

Example 39a: (1S,3S,4R,7S)-7-(p-Methoxybenzyloxy)-1-(p-methoxybenzyloxymethyl)-3-phenyl-2,5-dioxabicyclo[2.2.1]heptane [Compound **13a** of Scheme 2]

Cyclization of compound **12a** (540 mg, 1.13 mmol) in the presence of TMAD (310 mg, 1.8 mmol), PBU<sub>3</sub> (364 mg, 1.8 mmol) and benzene (10 cm<sup>3</sup>) followed by the general work-up procedure and column chromatography afforded compound **13a** as a colorless oil (400 mg, 77%); *R*<sub>f</sub> 0.51 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v);  $\delta_{\text{H}}$  (CDCl<sub>3</sub>) 7.36-7.33 (7 H, m), 7.10 (2 H, d, *J* 8.3), 6.88 (2 H, d, *J* 8.7), 6.78 (2 H, d, *J* 8.7), 5.17 (1 H, s, H-3), 4.59 (2 H, br s, -CH<sub>2</sub>(MPM)), 4.43 (1 H, d, *J* 11.3, -CH<sub>2</sub>(MPM)), 4.34 (1 H, d, *J* 11.3, -CH<sub>2</sub>(MPM)), 4.19 (1 H, s, H-4), 4.09 (1 H, d, *J* 7.7, H-6), 4.06 (1 H, d, *J* 7.7, H-6), 4.01 (1 H, s, H-7), 3.82-3.77 (5 H, m, -C<sub>1</sub>-CH<sub>2</sub>-O-, OCH<sub>3</sub>), 3.76 (3 H, s, -OCH<sub>3</sub>);  $\delta_{\text{C}}$  (CDCl<sub>3</sub>) 159.4, 159.3, 139.4 (C-1'), 130.3, 129.7, 129.5, 129.3, 128.5, 127.5, 125.4, 113.9, 113.8, 85.9 (C-1), 84.1 (C-3), 81.1 (C-4), 77.4 (C-7), 73.7 (-

$\text{CH}_2(\text{MPM})$ ), 73.4 (C-6), 71.8 ( $-\text{CH}_2(\text{MPM})$ ), 66.3 ( $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 55.4 ( $-\text{OCH}_3$ ), 55.3 ( $-\text{OCH}_3$ );  $m/z$  (FAB) 467  $[\text{M}+\text{Na}-\text{H}_2\text{O}]^+$ , 461  $[\text{M}-\text{H}]^+$ .

**Example 39b:** (1S,3S,4R,7S)-3-(4-Fluoro-3-methylphenyl)-7-(p-methoxybenzyloxy)-1-(p-methoxybenzyloxymethyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 13b of Scheme 2]

Cyclization of compound **12b** (550 mg, 1.08 mmol) in the presence of TMAD (275 mg, 1.6 mmol),  $\text{PBU}_3$  (325 mg, 1.6 mmol) and benzene ( $10\text{ cm}^3$ ) followed by the general work-up procedure and column chromatography afforded compound **13b** as a colorless oil (445 mg, 84%);  $R_f$  0.52 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  98:2, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 7.28 (2 H, d,  $J$  8.7), 7.11 (2 H, d,  $J$  8.6), 10 7.08-7.09 (2 H, m, H-2' and H-6'), 6.94 (1 H, dd,  $J$  8.5 and 9.2, H-5'), 6.88 (2 H, d,  $J$  8.6), 6.79 (2 H, d,  $J$  8.4), 5.08 (1 H, s, H-3), 4.62-4.55 (2 H, m,  $-\text{CH}_2(\text{MPM})$ ), 4.45 (1 H, d,  $J$  11.1,  $-\text{CH}_2(\text{MPM})$ ), 4.36 (1 H, d,  $J$  11.6,  $-\text{CH}_2(\text{MPM})$ ), 4.13 (1 H, s, H-4), 4.07, 4.03 (1 H each, 2d,  $J$  7.6 each, H-6), 3.99 (1 H, s, H-7), 3.81 (2 H, m,  $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 3.80 (3 H, s,  $-\text{OCH}_3$ ), 3.77 (3 H, s,  $-\text{OCH}_3$ ), 2.23 (3 H, d,  $J$  1.6, Ar- $\text{CH}_3$ );  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 162.3 (C-4'), 159.4, 159.3, 134.8, 134.7, 15 130.3, 129.6, 129.5, 129.2, 128.5, 128.4, 128.3, 124.2, 115.1, 114.8, 113.9, 113.8, 85.9 (C-1), 83.5 (C-3), 81.0 (C-4), 77.1 (C-7), 73.6 ( $-\text{CH}_2(\text{MPM})$ ), 73.4 (C-6), 71.8 ( $-\text{CH}_2(\text{MPM})$ ), 66.2 ( $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 55.4 ( $-\text{OCH}_3$ ), 55.3 ( $-\text{OCH}_3$ ), 14.7 (d,  $J$  3.3, Ar- $\text{CH}_3$ );  $m/z$  (FAB) 494  $[\text{M}]^+$ , 493  $[\text{M}-\text{H}]^+$ .

**Example 39c:** (1S,3S,4R,7S)-7-(p-Methoxybenzyloxy)-1-(p-methoxybenzyloxymethyl)-3-(1-naphthyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 13c of Scheme 2]

Cyclization of compound **12c** (700 mg, 1.32 mmol) in the presence of TMAD (345 mg, 2.0 mmol),  $\text{PBU}_3$  (405 mg, 2.0 mmol) and benzene ( $15\text{ cm}^3$ ) followed by the general work-up procedure and column chromatography afforded compound **13c** as a colorless oil (526 mg, 25 78%);  $R_f$  0.53 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  98:2, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 7.91-7.86 (2 H, m), 7.78 (1 H, d,  $J$  8.2), 7.73 (1 H, d,  $J$  7.1), 7.53-7.46 (3 H, m), 7.32 (2 H, d,  $J$  8.7), 7.04 (2 H, d,  $J$  8.7), 6.90 (2 H, d,  $J$  8.3), 6.71 (2 H, d,  $J$  8.6), 5.79 (1 H, s, H-3), 4.67-4.61 (2 H, m,  $-\text{CH}_2(\text{MPM})$ ), 4.43 (1 H, s, H-4), 4.38 (1 H, d,  $J$  11.2,  $-\text{CH}_2(\text{MPM})$ ), 4.27 (1 H, d,  $J$  10.9,  $-\text{CH}_2(\text{MPM})$ ), 4.16 (2 H, br s, H-6), 4.08 (1 H, s, H-7), 3.91, 3.87 (1 H each, 2d,  $J$  11.0 each,  $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 3.81 (3 H, s,  $-\text{OCH}_3$ ), 3.72 30 (3 H, s,  $-\text{OCH}_3$ );  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 159.3, 134.6 (C-1'), 133.5, 130.3, 129.8, 129.7, 129.4, 129.3, 128.9, 128.1, 126.4, 125.8, 125.6, 123.8, 122.7, 113.9, 113.7, 85.7 (C-1), 82.3 (C-3), 79.9 (C-4), 78.2 (C-7), 73.7 ( $-\text{OCH}_2(\text{MPM})$ ), 73.5 (C-6), 71.8 ( $-\text{OCH}_2(\text{MPM})$ ), 66.3 ( $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 55.4 ( $-\text{OCH}_3$ ), 55.3 ( $-\text{OCH}_3$ );  $m/z$  (FAB) 512  $[\text{M}]^+$ , 511  $[\text{M}-\text{H}]^+$ .

Example 39d: (1S,3S,4R,7S)-7-(p-Methoxybenzyloxy)-1-(p-methoxybenzyloxymethyl)-3-(1-pyrenyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 13d of Scheme 2]

Cyclization of compound **12d** (650 mg, 1.08 mmol) in the presence of TMAD (275 mg, 1.6 mmol),  $\text{PBU}_3$  (325 mg, 1.6 mmol) and benzene (10  $\text{cm}^3$ ) followed by the general work-up 5 procedure and column chromatography afforded compound **13d** as a pale yellow solid (496 mg, 79%);  $R_f$  0.53 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  98:2, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 8.29 (1 H, d,  $J$  8.2), 8.18-8.12 (5 H, m), 8.08-8.01 (2 H, m), 7.96 (1 H, d,  $J$  7.5), 7.35 (2 H, d,  $J$  8.5), 6.97 (2 H, d,  $J$  8.9), 6.92 (2 H, d,  $J$  8.8), 6.60 (2 H, d,  $J$  8.8), 6.09 (1 H, s, H-3), 4.71-4.65 (2 H, m,  $-\text{CH}_2(\text{MPM})$ ), 4.49 (1 H, s, H-4), 4.34 (1 H, d,  $J$  11.4,  $-\text{CH}_2(\text{MPM})$ ), 4.23 (1 H, d,  $J$  11.1,  $-\text{CH}_2(\text{MPM})$ ), 4.25 (1 H, d,  $J$  7.6, H-10), 4.21 (1 H, d,  $J$  7.8, H-6), 4.16 (1 H, s, H-7), 3.95-3.94 (2 H, m,  $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 3.81 (3 H, s,  $-\text{OCH}_3$ ), 3.59 (3 H, s,  $-\text{OCH}_3$ );  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 159.4, 159.3, 132.2 ( $\text{C}-1'$ ), 131.4, 130.8, 130.7, 130.4, 129.5, 129.4, 128.0, 127.5, 127.4, 126.9, 126.1, 125.6, 125.4, 124.9, 124.8, 124.7, 123.6, 122.0, 113.9, 113.7, 85.9 ( $\text{C}-1$ ), 82.7 ( $\text{C}-3$ ), 80.6 ( $\text{C}-4$ ), 77.9 ( $\text{C}-7$ ), 73.9 ( $-\text{OCH}_2(\text{MPM})$ ), 73.5 ( $\text{C}-6$ ), 71.8 ( $-\text{OCH}_2(\text{MPM})$ ), 66.3 ( $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 55.4 ( $-\text{OCH}_3$ ), 55.2 ( $-\text{OCH}_3$ );  $m/z$  (FAB) 587 15  $[\text{M}+\text{H}]^+$ , 586  $[\text{M}]^+$ .

Example 39e: (1S,3S,4R,7S)-7-(p-Methoxybenzyloxy)-1-(p-methoxybenzyloxymethyl)-3-(2,4,5-trimethylphenyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 13e of Scheme 2]

Cyclization of compound **12e** (550 mg, 1.05 mmol) in the presence of TMAD (275 mg, 1.6 mmol),  $\text{PBU}_3$  (325 mg, 1.6 mmol) and benzene (10  $\text{cm}^3$ ) followed by the general work-up 20 procedure and column chromatography afforded compound **13e** as a colorless oil (425 mg, 80%);  $R_f$  0.52 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  98:2, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3$ ) 7.30 (2 H, d,  $J$  9.0), 7.24 (1 H, s, H-6'), 7.13 (2 H, d,  $J$  8.9), 6.89 (1 H, s, H-3'), 6.88 (2 H, d,  $J$  8.8), 6.79 (2 H, d,  $J$  8.6), 5.18 (1 H, s, H-3), 4.64-4.57 (2 H, m,  $-\text{CH}_2(\text{MPM})$ ), 4.46 (1 H, d,  $J$  11.2,  $-\text{CH}_2(\text{MPM})$ ), 4.36 (1 H, d,  $J$  11.5,  $-\text{CH}_2(\text{MPM})$ ), 4.18 (1 H, s, H-4), 4.14 (1 H, s, H-7), 4.09 (1 H, d,  $J$  7.9, H-6), 4.04 (1 H, d,  $J$  7.7, H-6), 3.86 (2 H, s,  $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 3.80 (3 H, s,  $-\text{OCH}_3$ ), 3.76 (3 H, s,  $-\text{OCH}_3$ ), 2.21 (6 H, s, 2 x Ar- $\text{CH}_3$ ), 2.17 (3 H, s, Ar- $\text{CH}_3$ );  $\delta_{\text{C}}$  ( $\text{CDCl}_3$ ) 159.4, 159.3, 135.5 ( $\text{C}-1'$ ), 134.4, 134.0, 131.7, 131.3, 130.5, 129.9, 129.4, 129.2, 127.2, 113.9, 113.8, 85.6 ( $\text{C}-1$ ), 82.4 ( $\text{C}-3$ ), 79.4 ( $\text{C}-4$ ), 77.6 ( $\text{C}-7$ ), 73.5 ( $-\text{OCH}_2(\text{MPM})$ ), 73.4 ( $\text{C}-6$ ), 71.8 ( $-\text{OCH}_2(\text{MPM})$ ), 66.3 ( $-\text{C}_1-\text{CH}_2-\text{O}-$ ), 55.4 ( $-\text{OCH}_3$ ), 55.3 30 ( $-\text{OCH}_3$ ), 19.5 ( $-\text{CH}_3$ ), 19.3 ( $-\text{CH}_3$ ), 18.4 ( $-\text{CH}_3$ );  $m/z$  (FAB) 504  $[\text{M}]^+$ , 503  $[\text{M}-\text{H}]^+$ .

Example 40: General procedure for the oxidative removal of the p-methoxybenzyl groups to give Compounds 14a-e as shown in Scheme 2.

To a stirred solution of Compound **13a-e** in  $\text{CH}_2\text{Cl}_2$  (containing a small amount of  $\text{H}_2\text{O}$ ) at room temperature, was added 2,3-dichloro-5,6-dicyanoquinone (DDQ) which resulted in an immediate appearance of a deep greenish-black color which slowly faded into pale brownish-yellow. The reaction mixture was vigorously stirred at room temperature for 4 h. The precipitate was removed by filtration through a short pad of silica gel and washed with EtOAc. The combined filtrate was washed, successively, with saturated aqueous  $\text{NaHCO}_3$  (2 x 25  $\text{cm}^3$ ) and brine (25  $\text{cm}^3$ ). The separated organic phase was dried ( $\text{Na}_2\text{SO}_4$ ), filtered and evaporated to dryness under reduced pressure. The crude product obtained was purified by column chromatography [4-5% (v/v) MeOH in  $\text{CH}_2\text{Cl}_2$ ] to give compounds **14a-e**.

10

Example 40a: (1S,3S,4R,7S)-7-Hydroxy-1-hydroxymethyl-3-phenyl-2,5-dioxabicyclo[2.2.1]heptane [Compound 14a of Scheme 2]

Compound **13a** (400 mg, 0.86 mmol) was treated with DDQ (600 mg, 2.63 mmol) in a mixture of  $\text{CH}_2\text{Cl}_2$  (10  $\text{cm}^3$ ) and  $\text{H}_2\text{O}$  (0.5  $\text{cm}^3$ ). After the general work-up procedure and column chromatography, compound **14a** was obtained as a white solid material (128 mg, 66%);  $R_f$  0.30 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  9:1, v/v);  $\delta_H$  (( $\text{CD}_3$ )<sub>2</sub>CO/ $\text{CD}_3\text{OD}$ ; ( $\text{CD}_3$ )<sub>2</sub>CO was added to the compound followed by addition of  $\text{CD}_3\text{OD}$  until a clear solution appeared) 7.40-7.22 (5 H, m), 4.99 (1 H, s), 4.09 (1 H, s), 4.04 (1 H, s), 4.01 (1 H, d,  $J$  7.7), 3.86 (1 H, d,  $J$  7.7), 3.90 (2 H, br s), 3.77 (2 H, br s);  $\delta_C$  (( $\text{CD}_3$ )<sub>2</sub>CO/ $\text{CD}_3\text{OD}$ ; ( $\text{CD}_3$ )<sub>2</sub>CO was added to the compound followed by addition of  $\text{CD}_3\text{OD}$  until a clear solution appeared) 140.0, 128.2, 127.2, 125.4, 87.2, 83.7, 83.5, 72.3, 70.2, 58.4;  $m/z$  (FAB) 223  $[\text{M}+\text{H}]^+$ .

Example 40b: (1S,3S,4R,7S)-3-(4-Fluoro-3-methylphenyl)-7-hydroxy-1-hydroxymethyl-2,5-dioxabicyclo[2.2.1]heptane [Compound 14b of Scheme 2]

Compound **13b** (400 mg, 0.81 mmol) was treated with DDQ (570 mg, 2.5 mmol) in a mixture of  $\text{CH}_2\text{Cl}_2$  (10  $\text{cm}^3$ ) and  $\text{H}_2\text{O}$  (0.5  $\text{cm}^3$ ). After the general work-up procedure and column chromatography, compound **14b** was obtained as a white solid material (137 mg, 67%);  $R_f$  0.31 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  9:1, v/v);  $\delta_H$  ( $\text{CD}_3\text{OD}$ ) 7.23 (1 H, d,  $J$  8.1), 7.19 (1 H, m), 6.99 (1 H, dd,  $J$  8.5 and 9.3), 4.99 (1 H, s), 4.09 (1 H, s), 4.06 (1 H, s), 4.03 (1 H, d,  $J$  7.6), 3.93-3.91 (3 H, m), 2.25 (3 H, d,  $J$  1.4);  $\delta_C$  ( $\text{CD}_3\text{OD}$ ) 161.9 (d,  $J$  243.3), 136.4 (d,  $J$  3.4), 129.6 (d,  $J$  5.0), 126.1 (d,  $J$  22.8), 125.5 (d,  $J$  8.0), 115.7 (d,  $J$  22.9), 88.5, 85.0, 84.3, 73.5, 71.3, 59.4, 14.5 (d,  $J$  3.7);  $m/z$  (FAB) 255  $[\text{M}+\text{H}]^+$ .

Example 40c: (1S,3S,4R,7S)-7-Hydroxy-1-hydroxymethyl-3-(1-naphthyl)-2,5-dioxabicyclo-[2.2.1]heptane [Compound 14b of Scheme 2]

Compound 13c (475 mg, 0.93 mmol) was treated with DDQ (600 mg, 2.63 mmol) in a mixture of  $\text{CH}_2\text{Cl}_2$  (10  $\text{cm}^3$ ) and  $\text{H}_2\text{O}$  (0.5  $\text{cm}^3$ ). After the general work-up procedure and 5 column chromatography, compound 14c was obtained as a white solid material (170 mg, 67%);  $R_f$  0.31 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  9:1, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3/\text{CD}_3\text{OD}$ ;  $\text{CD}_3\text{OD}$  was added to the compound followed by addition of  $\text{CDCl}_3$  until a clear solution appeared) 7.94-7.86 (2 H, m), 7.80-7.74 (2 H, m), 7.55-7.46 (3 H, m), 5.74 (1 H, s), 4.56 (2 H, br s), 4.37 (1 H, s), 4.24 (1 H, s), 4.17-4.11 (2 H, m), 4.04 (2 H, br s);  $\delta_{\text{C}}$  ( $\text{CDCl}_3/\text{CD}_3\text{OD}$ ;  $\text{CD}_3\text{OD}$  was added to the compound followed by 10 addition of  $\text{CDCl}_3$  until a clear solution appeared) 134.7, 134.0, 130.2, 129.3, 128.6, 126.8, 126.2, 125.8, 123.8, 122.8, 87.4, 83.1, 82.2, 73.1, 71.5, 59.0;  $m/z$  (FAB) 273  $[\text{M}+\text{H}]^+$ , 272  $[\text{M}]^+$ .

Example 40d: (1S,3S,4R,7S)-7-Hydroxy-1-hydroxymethyl-3-(1-pyrenyl)-2,5-dioxabicyclo-[2.2.1]heptane [Compound 14d of Scheme 2]

15 Compound 13d (411 mg, 0.7 mmol) was treated with DDQ (570 mg, 2.5 mmol) in a mixture of  $\text{CH}_2\text{Cl}_2$  (10  $\text{cm}^3$ ) and  $\text{H}_2\text{O}$  (0.5  $\text{cm}^3$ ). After the general work-up procedure and column chromatography, compound 14d was obtained as a white solid material (182 mg, 75%);  $R_f$  0.32 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  9:1, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3/\text{CD}_3\text{OD}$ ;  $\text{CD}_3\text{OD}$  was added to the compound followed by addition of  $\text{CDCl}_3$  until a clear solution appeared) 8.32 (1 H, d,  $J$  7.8), 8.23-8.18 (5 20 H, m), 8.06 (2 H, br s), 8.01 (1 H, d,  $J$  7.6), 6.06 (1H, s), 4.47 (1 H, s), 4.36 (1 H, s), 4.27-4.18 (2 H, m), 4.10 (2 H, br s);  $\delta_{\text{C}}$  ( $\text{CDCl}_3/\text{CD}_3\text{OD}$ ) 132.2, 131.0, 128.5, 127.8, 127.3, 126.5, 125.9, 125.7, 125.1, 123.6, 122.1, 87.7, 83.7, 82.6, 73.1, 71.4, 58.9;  $m/z$  (FAB) 347  $[\text{M}+\text{H}]^+$ , 346  $[\text{M}]^+$ .

Example 40e: (1S,3S,4R,7S)-7-Hydroxy-1-hydroxymethyl-3-(2,4,5-trimethylphenyl)-2,5-25 dioxabicyclo[2.2.1]heptane [Compound 14e of Scheme 2]

Compound 13e (355 mg, 0.7 mmol) was treated with DDQ (570 mg, 2.5 mmol) in a mixture of  $\text{CH}_2\text{Cl}_2$  (10  $\text{cm}^3$ ) and  $\text{H}_2\text{O}$  (0.5  $\text{cm}^3$ ). After the general usual work-up procedure and column chromatography, compound 14e was obtained as a white solid material (120 mg, 65%);  $R_f$  0.31 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  9:1, v/v);  $\delta_{\text{H}}$  ( $\text{CDCl}_3/\text{CD}_3\text{OD}$ ;  $\text{CD}_3\text{OD}$  was added to the compound 30 followed by addition of  $\text{CDCl}_3$  until a clear solution appeared) 7.23 (1 H, s), 6.92 (1 H, s), 5.14 (1 H, s), 4.26 (1 H, s), 4.10 (1 H, s), 4.08, (1 H, d,  $J$  7.7), 4.00-3.95 (3 H, m), 2.23 (6 H, s), 2.21 (1 H, s);  $\delta_{\text{C}}$  ( $\text{CDCl}_3/\text{CD}_3\text{OD}$ ;  $\text{CD}_3\text{OD}$  was added to the compound followed by addition of  $\text{CDCl}_3$  until a clear solution appeared) 135.6, 133.9, 133.8, 131.7, 131.2, 126.6, 86.6, 82.1, 81.9, 72.3, 70.6, 58.5, 19.2, 19.0, 18.1;  $m/z$  (FAB) 265  $[\text{M}+\text{H}]^+$ , 264  $[\text{M}]^+$ .



Example 41: General procedure for dimethoxytritylation of compounds 14a-e to give Compounds 15a-e as shown in Scheme 2.

4,4'-Dimethoxytrityl chloride (DMTCI) was added in one portion to a stirred solution of 5 compound 14a-e in anhydrous pyridine. After stirring the mixture at room temperature for 4 h, methanol (0.2 cm<sup>3</sup>) was added and the resulting mixture was evaporated to dryness under reduced pressure. The residue was coevaporated with anhydrous CH<sub>3</sub>CN (2 x 5 cm<sup>3</sup>) and anhydrous toluene (2 x 5 cm<sup>3</sup>) and then dissolved in CH<sub>2</sub>Cl<sub>2</sub> (20 cm<sup>3</sup>, traces of acid removed by filtration through a short pad of basic alumina). The resulting solution was washed, 10 successively, with saturated aqueous NaHCO<sub>3</sub> (2 x 10 cm<sup>3</sup>) and brine (10 cm<sup>3</sup>). The separated organic phase was dried (Na<sub>2</sub>SO<sub>4</sub>), filtered and evaporated to dryness under reduced pressure. The crude product obtained was purified by column chromatography [0.25- 0.50% (v/v) MeOH in CH<sub>2</sub>Cl<sub>2</sub>, containing 0.5% Et<sub>3</sub>N] affording compounds 15a-e.

15 Example 41a: (1R,3S,4R,7S)-1-(4,4'-Dimethoxytrityloxymethyl)-7-hydroxy-3-phenyl-2,5-dioxabicyclo[2.2.1]heptane [Compound 15a of Scheme 2]

Dimethoxytritylation of compound 14a (108 mg, 0.49 mmol) using DMTCI (214 mg, 0.63 mmol) in anhydrous pyridine (2 cm<sup>3</sup>) followed by the general work-up procedure and column chromatography afforded compound 15a as a white solid material (180 mg, 71%); *R*<sub>f</sub> 20 0.31 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v); δ<sub>H</sub> (CDCl<sub>3</sub>) 7.66-7.21 (14 H, m), 6.84 (4 H, d, *J* 8.8), 5.19 (1 H, s), 4.29 (1 H, s), 4.13 (1 H, s), 4.07 (1 H, d, *J* 8.4), 4.01 (1 H, d, *J* 8.3), 3.78 (6 H, s), 3.55 (1 H, d, *J* 10.2), 3.50 (1 H, d, *J* 10.7), 2.73 (1 H, br s); δ<sub>C</sub> (CDCl<sub>3</sub>) 158.6, 149.8, 144.9, 139.4, 136.2, 135.9, 135.8, 130.3, 130.2, 128.5, 128.3, 128.0, 127.6, 126.9, 125.4, 123.9, 113.3, 86.4, 86.0, 83.8, 83.4, 73.0, 71.6, 60.2, 55.3; *m/z* (FAB) 525 [M+H]<sup>+</sup>, 524 [M]<sup>+</sup>.

25

Example 41b: (1R,3S,4R,7S)-1-(4,4'-Dimethoxytrityloxymethyl)-3-(4-fluoro-3-methylphenyl)-7-hydroxy-2,5-dioxabicyclo[2.2.1]heptane [Compound 15b of Scheme 2]

Dimethoxytritylation of compound 14b (95 mg, 0.38 mmol) using DMTCI (129 mg, 0.42 mmol) in anhydrous pyridine (2 cm<sup>3</sup>) followed by the general work-up procedure and 30 column chromatography afforded compound 15b as a white solid material (126 mg, 61%); *R*<sub>f</sub> 0.32 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v); δ<sub>H</sub> (CDCl<sub>3</sub>) 7.53-7.15 (11 H, m), 6.97 (1 H, dd, *J* 8.7 and 8.9), 6.84 (4 H, d, *J* 8.8), 5.11 (1 H, s), 4.26 (1 H, d, *J* 3.9), 4.08 (1 H, s), 4.03 (1 H, d, *J* 8.0), 3.95 (1 H, d, *J* 8.0), 3.78 (6 H, s), 3.54 (1 H, d, *J* 10.5), 3.47 (1 H, d, *J* 10.1), 2.26 (3 H, d, *J* 1.5), 2.08 (1 H, br s); δ<sub>C</sub> (CDCl<sub>3</sub>) 160.8 (d, *J* 244.1), 158.7, 144.9, 135.9, 134.7, 134.6, 130.3, 130.2,

130.1, 128.5, 128.4, 128.3, 128.0, 127.0, 125.2, 124.9, 124.4, 124.3, 115.2, 114.9, 113.4, 86.5, 86.0, 83.7, 83.0, 72.9, 71.7, 60.1, 55.3, 14.8 (d,  $J$  3.1);  $m/z$  (FAB) 556  $[M]^+$ .

Example 41c: (1R,3S,4R,7S)-1-(4,4'-Dimethoxytrityloxymethyl)-7-hydroxy-3-(1-naphthyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 15c of Scheme 2]

Dimethoxytritylation of compound **14c** (125 mg, 0.46 mmol) using DMTCI (170 mg, 0.5 mmol) in anhydrous pyridine (2 cm<sup>3</sup>) followed by the general work-up procedure and column chromatography afforded compound **15c** as a white solid material (158 mg, 60%);  $R_f$  0.35 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v);  $\delta_H$  (CDCl<sub>3</sub>) 7.95-7.86 (3 H, m), 7.79 (1 H, d,  $J$  8.3), 7.58-7.41 10 (9 H, m), 7.35-7.23 (3 H, m), 6.86 (4 H, d,  $J$  8.8), 5.80 (1 H, s), 4.36 (1 H, s), 4.32 (1 H, d,  $J$  6.5), 4.17 (1 H, d,  $J$  8.3), 4.06 (1 H, d,  $J$  8.0), 3.78 (6 H, s), 3.62-3.56 (2 H, m), 2.00 (1 H, d,  $J$  6.6);  $\delta_C$  (CDCl<sub>3</sub>) 158.7, 144.9, 136.0, 135.9, 134.5, 133.6, 130.3, 129.8, 129.0, 128.3, 128.2, 128.1, 127.0, 126.5, 125.9, 125.6, 123.9, 122.6, 113.4, 86.6, 85.7, 82.5, 81.7, 73.1, 72.6, 60.2, 55.3;  $m/z$  (FAB) 575  $[M+H]^+$ , 574  $[M]^+$ .

15

Example 41d: (1R,3S,4R,7S)-1-(4,4'-Dimethoxytrityloxymethyl)-7-hydroxy-3-(1-pyrenyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 15d of Scheme 2]

Dimethoxytritylation of the compound **14d** (130 mg, 0.38 mmol) using DMTCI (140 mg, 0.42 mmol) in anhydrous pyridine (2 cm<sup>3</sup>) followed by the general work-up procedure and column chromatography afforded compound **15d** as a white solid material (147 mg, 61%);  $R_f$  0.37 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v);  $\delta_H$  (CDCl<sub>3</sub>) 8.46 (1 H, d,  $J$  8.0), 8.19-8.00 (7 H, m), 7.61 (2 H, dd,  $J$  1.6 and 7.4), 7.48 (4 H, d,  $J$  8.3), 7.35 (2 H, dd,  $J$  7.2 and 7.5), 7.25 (1 H, m), 7.15 (1 H, m), 6.88 (4 H, d,  $J$  9.0), 6.10 (1 H, s), 4.46 (1 H, s), 4.43 (1 H, br s), 4.25 (1 H, d,  $J$  8.1), 4.12 (1 H, d,  $J$  8.1), 3.79 (6H, s), 3.71-3.63 (2 H, m), 2.22 (1 H, br s);  $\delta_C$  (CDCl<sub>3</sub>) 158.7, 149.8, 144.9, 136.1, 136.0, 135.9, 132.1, 131.4, 130.9, 130.6, 130.3, 130.2, 129.2, 129.1, 128.4, 128.3, 128.2, 128.1, 127.5, 127.4, 127.0, 126.9, 126.2, 125.5, 125.4, 124.9, 124.8, 124.7, 123.8, 123.7, 121.9, 113.4, 86.6, 86.1, 83.2, 82.2, 73.2, 72.4, 60.3, 55.3;  $m/z$  (FAB) 649  $[M+H]^+$ , 648  $[M]^+$ .

Example 41e: (1R,3S,4R,7S)-1-(4,4'-Dimethoxytrityloxymethyl)-7-hydroxy-3-(2,4,5-trimethylphenyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 15e of Scheme 2]

Dimethoxytritylation of compound **14e** (80 mg, 0.3 mmol) using DMTCI (113 mg, 0.33 mmol) in anhydrous pyridine (2 cm<sup>3</sup>) followed by the general work-up procedure and column chromatography afforded compound **15e** as a white solid material (134 mg, 78%);  $R_f$  0.32 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v);  $\delta_H$  (CDCl<sub>3</sub>) 7.55 (2 H, d,  $J$  7.9), 7.45-7.42 (4 H, m), 7.32-7.21 (4 H,

m), 6.93 (1 H, s), 6.84 (4 H, d,  $J$  8.2), 5.20 (1 H, s), 4.40 (1 H, s), 4.08 (1 H, s), 4.04 (1 H, d,  $J$  8.3), 3.95 (1 H, d,  $J$  8.2), 3.78 (6 H, s), 3.56 (1 H, d,  $J$  10.5), 3.47 (1 H, d,  $J$  10.2), 2.24 (3 H, s), 2.22 (3 H, s), 2.19 (3 H, s);  $\delta$  ( $\text{CDCl}_3$ ) 158.6, 145.0, 136.0, 135.7, 134.4, 134.2, 131.8, 131.3, 130.3, 130.2, 128.3, 128.0, 127.2, 126.9, 113.3, 86.4, 85.7, 82.1, 81.8, 73.0, 71.8, 60.2, 55.3, 19.6, 19.3, 18.4;  $m/z$  (FAB) 567  $[\text{M}+\text{H}]^+$ , 566  $[\text{M}]^+$ .

Example 42: General procedure for synthesis of the phosphoramidite derivatives 16a-e as shown in Scheme 2.

2-Cyanoethyl  $N,N'$ -diisopropylphosphoramidochloridite was added dropwise to a stirred 10 solution of nucleoside 15a-e and  $N,N'$ -diisopropylethylamine (DIPEA) in anhydrous  $\text{CH}_2\text{Cl}_2$  at room temperature. After stirring the mixture at room temperature for 6 h, methanol ( $0.2 \text{ cm}^3$ ) was added and the resulting mixture diluted with EtOAc ( $20 \text{ cm}^3$ , containing 0.5%  $\text{Et}_3\text{N}$ , v/v). The organic phase was washed, successively, with saturated a.  $\text{NaHCO}_3$  ( $2 \times 10 \text{ cm}^3$ ) and brine ( $10 \text{ cm}^3$ ). The separated organic phase was dried ( $\text{Na}_2\text{SO}_4$ ), filtered and evaporated to dryness 15 under reduced pressure. The residue obtained was purified by column chromatography [25-30% (v/v) EtOAc in  $n$ -hexane containing 0.5%  $\text{Et}_3\text{N}$ ] to give the amidites 16a-e.

Example 42a: Synthesis of (1R,3S,4R,7S)-7-[2-Cyanoethoxy(diisopropylamino) phosphinoxy]-1-(4,4'-dimethoxytrityloxymethyl)-3-phenyl-2,5-dioxabicyclo[2.2.1] heptane 20 [Compound 16a of Scheme 2]

Treatment of compound 15a (170 mg, 0.32 mmol) with 2-cyanoethyl  $N,N'$ -diisopropylphosphoramidochloridite (85 mg, 0.36 mmol) in the presence of DIPEA ( $0.4 \text{ cm}^3$ ) and anhydrous  $\text{CH}_2\text{Cl}_2$  ( $2.0 \text{ cm}^3$ ) followed by the general work-up procedure and column chromatography afforded phosphoramidite 16a as a white solid material (155 mg, 66%);  $R_f$  25 0.45, 0.41 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  98:2, v/v);  $\delta$  ( $\text{CDCl}_3$ ) 149.3, 148.9.

Example 42b: (1R,3S,4R,7S)-7-[2-Cyanoethoxy(diisopropylamino) phosphinoxy]-1-(4,4'-dimethoxytrityloxymethyl)-3-(4-fluoro-3-methylphenyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 16b of Scheme 2]

30 Treatment of compound 15b (95 mg, 0.17 mmol) with 2-cyanoethyl  $N,N'$ -diisopropylphosphoramidochloridite (53 mg, 0.22 mmol) in the presence of DIPEA ( $0.3 \text{ cm}^3$ ) and anhydrous  $\text{CH}_2\text{Cl}_2$  ( $2.0 \text{ cm}^3$ ) followed by the general work-up procedure and column chromatography afforded phosphoramidite 16b as a white solid material (85 mg, 66%);  $R_f$  0.45, 0.41 ( $\text{CH}_2\text{Cl}_2/\text{MeOH}$  98:2, v/v);  $\delta$  ( $\text{CDCl}_3$ ) 149.3, 148.8.

Example 42c: Synthesis of (1R,3S,4R,7S)-7-[2-Cyanoethoxy(diisopropylamino)phosphinoxy]-1-(4,4'-dimethoxytrityloxymethyl)-3-(1-naphthyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 16c of Scheme 2]

Treatment of compound **5c** (158 mg, 0.28 mmol) with 2-cyanoethyl *N,N'*-diisopropylphosphoramidochloridite (75.7 mg, 0.32 mmol) in the presence of DIPEA (0.4 cm<sup>3</sup>) and anhydrous CH<sub>2</sub>Cl<sub>2</sub> (2.0 cm<sup>3</sup>) followed by the general work-up procedure and column chromatography afforded phosphoramidite **16c** as a white solid material (127 mg, 60%); *R<sub>f</sub>* 0.47, 0.44 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v); δ<sub>p</sub> (CDCl<sub>3</sub>) 149.2, 149.1.

10 Example 42d: Synthesis of (1R,3S,4R,7S)-7-[2-Cyanoethoxy(diisopropylamino)phosphinoxy]-1-(4,4'-dimethoxytrityloxymethyl)-3-(1-pyrenyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 16d of Scheme 2]

Treatment of compound **15d** (140 mg, 0.22 mmol) with 2-cyanoethyl *N,N'*-diisopropylphosphoramidochloridite (64 mg, 0.27 mmol) in the presence of DIPEA (0.3 cm<sup>3</sup>) and anhydrous CH<sub>2</sub>Cl<sub>2</sub> (2.0 cm<sup>3</sup>) followed by the general work-up procedure and column chromatography afforded phosphoramidite **16d** as a white solid material (124 mg, 68%); *R<sub>f</sub>* 0.51, 0.47 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v); δ<sub>p</sub> (CDCl<sub>3</sub>) 149.4, 149.1.

Example 42e: Synthesis of (1R,3S,4R,7S)-7-[2-Cyanoethoxy(diisopropylamino)phosphinoxy]-1-(4,4'-dimethoxytrityloxymethyl)-3-(2,4,5-trimethylphenyl)-2,5-dioxabicyclo[2.2.1]heptane [Compound 16e of Scheme 2]

Treatment of compound **15e** (130 mg, 0.23 mmol) with 2-cyanoethyl *N,N'*-diisopropylphosphoramidochloridite (64 mg, 0.27 mmol) in the presence of DIPEA (0.3 cm<sup>3</sup>) and anhydrous CH<sub>2</sub>Cl<sub>2</sub> (2.0 cm<sup>3</sup>) followed by the general work-up procedure and column chromatography afforded phosphoramidite **16e** as a white solid material (111 mg, 63%); *R<sub>f</sub>* 0.44, 0.42 (CH<sub>2</sub>Cl<sub>2</sub>/MeOH 98:2, v/v); δ<sub>p</sub> (CDCl<sub>3</sub>) 149.0.

#### Example 43: Synthesis, deprotection and purification of oligonucleotides

All oligomers were prepared using the phosphoramidite approach on a Biosearch 8750 DNA synthesizer in 0.2 μmol scale on CPG solid supports (BioGenex). The stepwise coupling efficiencies for phosphoramidites **16a-c** (10 min coupling time) and phosphoramidites **16d** and **16e** (20 min coupling time) were >96% and for unmodified deoxynucleoside and ribonucleoside phosphoramidites (with standard coupling time) generally >99%, in all cases using 1*H*-tetrazole as activator. After standard deprotection and cleavage from the solid support

using 32% aqueous ammonia (12 h, 55 °C), the oligomers were purified by precipitation from ethanol. The composition of the oligomers were verified by MALDI-MS analysis and the purity (>80%) by capillary gel electrophoresis. Selected MALDI-MS data ( $[M-H]^+$ ; found/calcd.: ON3 2731/2733; ON4 2857/2857; ON6 3094/3093).

5

#### Example 44. Thermal denaturation studies

The thermal denaturation experiments were performed on a Perkin-Elmer UV/VIS spectrometer fitted with a PTP-6 Peltier temperature-programming element using a medium salt buffer solution (10 mM sodium phosphate, 100 mM sodium chloride, 0.1 mM EDTA, pH 7.0).

10 Concentrations of 1.5 mM of the two complementary strands were used assuming identical extinction coefficients for modified and unmodified oligonucleotides. The absorbance was monitored at 260 nm while raising the temperature at a rate of 1 °C per min. The melting temperatures ( $T_m$  values) of the duplexes were determined as the maximum of the first derivatives of the melting curves obtained.

15

#### Example 45: Synthesis of compounds 16a-16e and oligomers containing monomers 17a-17e

LNA containing the derivatives **17a-17e** (Figure 1, Scheme 1, Scheme 2), were synthesized, all based on the LNA-type 2'-O,4'-C-methylene- $\beta$ -D-ribofuranosyl moiety which is known to adopt a locked C3'-endo RNA-like furanose conformation [S. Obika, D. Nanbu, Y. Hari, K. Morio, Y. In, T. Ishida, and T. Imanishi, *Tetrahedron Lett.*, 1997, **38**, 8735; S. K. Singh, P. Nielsen, A. A. Koshkin and J. Wengel, *Chem. Commun.*, 1998, 455; A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and J. Wengel, *Tetrahedron*, 1998, **54**, 3607; S. Obika, D. Nanbu, Y. Hari, J. Andoh, K. Morio, T. Doi and T. Imanishi, *Tetrahedron Lett.*, 1998, **39**, 5401]. The syntheses of the phosphoramidite building blocks **16a-16e** suitable for incorporation of the LNA-type aryl C-glycosides **17a-17e** are shown in Scheme 1 and Scheme 2 and described in details in the experimental section. In the design of an appropriate synthetic route, it was decided to utilize a reaction similar to one described recently in the literature. Thus, stereoselective attack of Grignard reagents of various heterocycles on a carbonyl group of an aldehyde corresponding to aldehyde **11** (Scheme 2) but 20 with two *O*-benzyl groups instead of the two *p*-methoxybenzyl groups of aldehyde **11** (Scheme 2) has been reported to furnish locked-C-nucleosides [S. Obika, Y. Hari, K. Morio and T. Imanishi, *Tetrahedron Lett.*, 2000, **41**, 215; S. Obika, Y. Hari, K. Morio and T. Imanishi, *Tetrahedron Lett.*, 2000, **41**, 221]. The key intermediate in the synthetic route selected herein, namely the novel aldehyde **11** was synthesized from the known furanoside **1** [R. Yamaguchi, T. Imanishi, S. Kohgo, H. Horie and H. Ohru, *Biosci. Biotechnol. Biochem.*, 1999, **63**, 736]

35

following two different routes. In general, *O*-(*p*-Methoxy)benzyl protection was desirable instead of *O*-benzyl protection as removal of the benzyl protection at a later stage (*i.e.* 13—>14) could also likely result in the cleavage of the benzylic O-C<sub>1</sub> bond present, *e.g.*, in compounds 13 and 14 (Scheme 2). In one route to give aldehyde 11, regioselective *p*-5 methoxybenzylation of the furanoside 1, followed by mesylation and methanolysis yielded the anomeric mixture of the methyl furanosides 9. Base induced cyclization followed by acetyl hydrolysis afforded the aldehyde 11 in approximately 24% overall yield from 1 (Scheme 1 and Scheme 2). This yield was improved to following a different strategy. Thus, di-*O*-mesylation of 1 followed by methanolysis and base induced intramolecular nucleophilic attack from the 2-OH 10 group afforded the cyclized anomeric mixture of methyl furanoside 4. Substitution of the remaining mesyloxy group of 4 with an acetate group, followed by deacetylation, *p*-methoxybenzylation and then acetyl hydrolysis afforded the required aldehyde 11 (Scheme 1).

Coupling of the aldehyde 11 with different aryl Grignard reagents yielded selectively one epimer of each of the compounds 12a-e in good yields (see experimental section for further 15 details on this and other synthetic steps). Each of the diols 12a-e was cyclized under Mitsunobu conditions (TMAD, PBU<sub>3</sub>) to afford the bicyclic  $\beta$ -C-nucleoside derivatives 13a-e. Oxidative removal of the *p*-methoxybenzyl protections was achieved in satisfactory yields using DDQ. Subsequent, selective 4,4'-dimethoxytritylation ( to give compounds 15a-e) followed by phosphorylation afforded the phosphoramidite building blocks 16a-e in satisfactory yields. The 20 configuration of compounds 13, and thus also compounds 11, 12 and 14-17 were assigned based on <sup>1</sup>H NMR spectroscopy, including NOE experiments.

All oligomers were prepared in the 0.2  $\mu$ mol scale using the phosphoramidite approach. The stepwise coupling efficiencies for phosphoramidites 16a-c (10 min coupling time) and phosphoramidites 16d and 16e (20 min coupling time) were >96% and for unmodified 25 deoxynucleoside and ribonucleoside phosphoramidites (with standard coupling time) generally >99%, in all cases using 1*H*-tetrazole as activator. After standard deprotection and cleavage from the solid support using 32% aqueous ammonia (12 h, 55 °C), the oligomers were purified by precipitation from ethanol. The composition of the oligomers were verified by MALDI-MS analysis and the purity (>80%) by capillary gel electrophoresis.

30

#### Example 46 Thermal denaturation studies to evaluate hybridization properties

The hybridization of the oligonucleotides ON1-ON11 (Table 8 below) toward four 9-mer DNA targets with the central base being each of four natural bases were studied by thermal denaturation experiments (*T<sub>m</sub>* measurements; see the experimental section for details). 35 Compared to the DNA reference ON1, introduction of one abasic LNA monomer Ab<sup>L</sup> (ON2)

has earlier been reported to prevent the formation of a stable duplex above 0 °C (only evaluated with adenine as the opposite base) [L. Kværnø and J. Wengel, *Chem. Commun.*, 1999, 657]. With the phenyl monomer **17a** (**ON3**),  $T_m$  values in the range of 5-12 °C was observed. Thus, the phenyl moiety stabilizes the duplexes compared to **Ab<sup>L</sup>**, but universal hybridization is not achieved as a preference for a central adenine base in the complementary target strand is indicated (Table 8). In addition, significant destabilization compared to the **ON1**:DNA reference duplex was observed. Results similar to those obtained for **ON3** were obtained for oligomers isosequential with **ON3** but containing **17b**, **17c** or **17e** instead of **17a** as the central monomer (Table 8, **ON7**, **ON8** and **ON9**, respectively).

10

Table 8. Thermal denaturation experiments ( $T_m$  values shown) for **ON1**-**ON11** towards DNA complements with each of the four natural bases in the central position<sup>a</sup>

DNA target: 3'-d(CACTYTACG)		Y:	A	C	G	T
15	<b>ON1</b> 5'-d(GTGATATGC)		28	11	12	19
	<b>ON2</b> 5'-d(GTGAAb <sup>L</sup> ATGC)		<3	n.d.	n.d.	n.d.
	<b>ON3</b> 5'-d(GTGA <b>17a</b> ATGC)		12	5	6	7
	<b>ON4</b> 5'-d(GTGA <b>17d</b> ATGC)		18	17	18	19
	<b>ON5</b> 5'-d[2'-OMe(GTGATATGC)]		35	14	19	21
20	<b>ON6</b> 5'-d[2'-OMe(GT <sup>L</sup> GA <b>17d</b> AT <sup>L</sup> GC)]		39	38	37	40
	<b>ON7</b> 5'-d(GTGA <b>17b</b> ATGC)		15	7	6	8
	<b>ON8</b> 5'-d(GTGA <b>17c</b> ATGC)		15	7	6	9
	<b>ON9</b> 5'-d(GTGA <b>17e</b> ATGC)		13	6	6	7
	<b>ON10</b> 5'-d[2'-OMe(GT <sup>L</sup> GA <b>17b</b> AT <sup>L</sup> GC)]		31	25	26	27
25	<b>ON11</b> 5'-d[2'-OMe(GT <sup>L</sup> GA <b>17c</b> AT <sup>L</sup> GC)]		34	27	27	32

<sup>a</sup> Melting temperatures ( $T_m$  values/°C) measured as the maximum of the first derivative of the melting curve ( $A_{260}$  vs temperature) recorded in medium salt buffer (10 mM sodium phosphate, 100 mM sodium chloride, 0.1 mM EDTA, pH 7.0) using 1.5  $\mu$ M concentrations of the two strands; A = adenine monomer, C = cytosine monomer, G = guanine monomer, T = thymine monomer; See Figure 1 and/or Scheme 2 for structures of **T<sup>L</sup>**, **Ab<sup>L</sup>** and **17a-17e**; DNA sequences are shown as d(sequence) and 2'-OMe-RNA sequences as 2'-OMe(sequence); "n.d." denotes "not determined". The data reported for **ON1** have been reported earlier [A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and J. Wengel, *Tetrahedron*, 1998, 54, 3607]. The data reported for **ON2** has been reported earlier [L. Kværnø and J. Wengel, *Chem. Commun.*, 1999, 657].

35

The pyrene LNA nucleotide **17d** (in **ON4**) displays more encouraging properties (Table 8). Firstly, the binding affinity towards all four complements is increased compared to **ON3**

(containing **17a**). Secondly, universal hybridization is observed as shown by the four  $T_m$  values all being within 17-19 °C. With respect to universal hybridization, **17d** thus parallels the pyrene DNA derivative **Py** [T. J. Matray and E. T. Kool, *J. Am. Chem. Soc.*, 1998, **120**, 6191], but the decrease in thermal stability compared to the **ON1**:DNA reference is more pronounced for **17d** 5 (~10 °C) than reported for **Py** (~5 °C in a 12-mer polypyrimidine DNA sequence) [T. J. Matray and E. T. Kool, *J. Am. Chem. Soc.*, 1998, **120**, 6191]. It therefore appears that stacking (or intercalation) by the pyrene moiety is not favored by the conformational restriction of the furanose ring of **17d**, although comparison of the thermal stabilities of **ON2**, **ON3** and **ON4** strongly indicate interaction of the pyrene moiety within the helix.

10 When measured against an RNA target [3'-r(CACUAUACG)], the  $T_m$  values (using identical experimental conditions as for the experiments described above) of **ON3** was 11.9 °C and of **ON4** was 12.7 °C. For oligomers **ON7**, **ON8** and **ON9** (Table 8), the corresponding  $T_m$  values were 11.7, 8.8 and 10.2 °C, respectively.

15 Example 47: The effect of pyrene LNA monomers in an RNA-like strand.

**ON5**, **ON6**, **ON10** and **ON11** (see Table 8 above), were synthesized. The former being composed entirely of 2'-OMe-RNA monomers and the latter three of six 2'-OMe-RNA monomers (see Figure 1), two LNA thymine monomers **T<sup>L</sup>** (see Figure 1), and one central LNA pyrene monomer **17d** (oligomer **ON6**), or one central monomer **17b** (**ON10**) or **17c** (**ON11**). A 20 sequence corresponding to **ON6** but with three **T<sup>L</sup>** monomers has earlier been shown to form a duplex with complementary DNA of very high thermal stability. **ON6** is therefore suitable for evaluation of the effect of introducing high-affinity monomers around a universal base. As seen in Table 8, the 2'-OMe-RNA reference **ON5** binds to the DNA complement with slightly increased thermal stability and conserved Watson-Crick discrimination (compared to the DNA 25 reference **ON1**). Indeed, the LNA/2'-OMe-RNA chimera **ON6** displays universal hybridization behavior as revealed from the four  $T_m$  values (37, 38, 39 and 40 °C). All four  $T_m$  values obtained for **ON6** are higher than the  $T_m$  values obtained for the two fully complementary reference duplexes **ON1**:DNA ( $T_m$  = 28 °C) and **ON5**:DNA ( $T_m$  = 35 °C).

These novel data demonstrate that the pyrene LNA monomer **17d** display universal 30 hybridization behavior both in a DNA context (**ON4**) and in an RNA-like context (**ON6**), and that the problem of decreased affinity of universal hybridization probes can be solved by the introduction of high-affinity monomers, e.g. 2'-OMe-RNA and/or LNA monomers. Increased affinities compared to **ON7** and **ON8** were obtained for **ON10** and **ON11**, respectively, but universal hybridization behavior was not obtained as a preference for a central adenine base in 35 the complementary target strand is indicated (Table 8 above).



Example 48: Base-pairing selectivity in hybridization probes.

A systematic thermal denaturation study with **ON6** (Table 11) was performed to determine base-pairing selectivity. For each of the four DNA complements (DNA target 5 strands; monomer **Y** = A, C, G or T) used in the study shown in Table 8 above, **ON6**, containing a central pyrene LNA monomer **17d**, was hybridized with all four base combinations in the neighboring position towards the 3'-end of **ON6** (DNA target strands; monomer **Z** = A, C, G or T, monomer **X** = T) and the same towards the 5'-end of **ON6** (DNA target strands; monomer **X** = A, C, G or T, monomer **Z** = T). In all eight subsets of four data points, 10 satisfactory to excellent Watson-Crick discrimination was observed between the match and the three mismatches (Table 11 below,  $\Delta T_m$  values in the range of 5-25 °C).

Table 11. Thermal denaturation experiments ( $T_m$  values shown) to evaluate the base-pairing selectivity of the bases neighboring the universal pyrene LNA monomer **17d** in the 2'-OMe-15 RNA/LNA chimera **ON6**. In the target strand [3'-d(CAC-XYZ-ACG)], the central three bases **XYZ** are varied among each of the four natural bases<sup>a</sup>

5'-[2'-OMe(GT <sup>L</sup> G-A17dA-T <sup>L</sup> GC)]							
3'-d(CAC -X Y Z- ACG)							
20							
XYZ	$T_m$ °C	XYZ	$T_m$ °C	XYZ	$T_m$ °C	XYZ	$T_m$ °C
TAA	26	TCA	22	TGA	22	TTA	29
TAC	26	TCC	29	TGC	26	TTG	31
TAG	24	TCG	24	TGG	30	TTC	32
25 TAT	<b>39</b>	TCT	<b>38</b>	TGT	<b>37</b>	TTT	<b>40</b>
AAT	18	ACT	27	AGT	22	ATT	28
CAT	30	CCT	31	CGT	27	CTT	35
GAT	14	GCT	28	GGT	16	GTT	27
TAT	<b>39</b>	TCT	<b>38</b>	TGT	<b>37</b>	TTT	<b>40</b>

30

<sup>a</sup> See caption below Table 8 for abbreviations and conditions used; The data for matched neighboring bases (**X** = **Z** = T) are shown in bold.

The results reported herein have several important implications for the design of probes 35 for universal hybridization: (1) Universal hybridization is possible with a conformationally

restricted monomer as demonstrated for the pyrene LNA monomer; (2) Universal hybridization behavior is feasible in an RNA context; (3) The binding affinity of probes for universal hybridization can be increased by the introduction of high-affinity monomers without compromising the universal hybridization and the base-pairing selectivity of bases neighboring 5 the universal base.

Based on the results reported herein, that chimeric oligonucleotides comprising pyrene and other known universal bases attached at various backbones (*e.g.* LNA-type monomers, ribofuranose monomers or deoxyribose monomers in 2'-OMe-RNA/LNA chimeric oligos) likewise will display attractive properties with respect to universal hybridization behavior. For 10 example, an oligomer identical with the 2'-OMe-RNA/LNA oligo ON6 but with the 17d monomer substituted by a pyrenyl-2'-OMe-ribonucleotide monomer.

#### Example 49: Chimeric oligonucleotides

These chimeric oligonucleotides are comprised of pyrene and other known universal 15 bases attached at various backbones (*e.g.* LNA-type monomers, ribofuranose monomers or deoxyribose monomers in 2'-OMe-RNA/LNA oligos). Experimentation with these chimeric oligonucleotides are for evaluating the possibility of obtaining similar results to the 2'-OMe-RNA/LNA oligo ON6 at a lower cost, for example, by substituting  $\text{Py}^{\text{L}}$  with a pyrenyl-2'-OMe-ribonucleotide monomer.

20

#### Example 50: The use of LNA Oligonucleotide Microarrays Provides Superior Sensitivity and Specificity in Expression Profiling

##### A. *In vitro* synthesis of the yeast spike RNAs

Amplification of the yeast genes was carried out by standard two-step PCR using yeast 25 genomic DNA as template (21). In the second PCR, a poly-T<sub>20</sub> tail was inserted in the amplicon. The DNA fragments were ligated into the pTRIamp18 vector (Ambion, USA) using the Quick Ligation Kit (New England Biolabs, USA) according to the manufacturers instructions and transformed into *E. coli* DH-5 $\alpha$  (21). Synthesis of *in vitro* RNA was done using the MEGAscript™ T7 Kit (Ambion, USA) according to the manufacturers instructions.

##### 30 B. Design of the LNA expression arrays

Capture probes were designed using the OligoDesign™ software as described in the previous examples.

##### C. Printing and hybridisation of the LNA expression microarrays

The LNA oligonucleotide microarrays were printed onto Immobilizer™ MicroArray Slides (Exiqon, Denmark) using the Packard Biochip I Arrayer (Packard, USA), with a spot volume of 2 x 300 pl of a 10  $\mu$ M capture probe solution. Four replicas of each capture probe were printed on each slide. Mixed staged *Caenorhabditis elegans* worm cultures were 5 cultivated according to standard protocols. RNA was extracted from worm samples using the FastRNA Kit, GREEN (Q-BIO, USA) according to the supplier's instructions. Fluorochrome-labelled first strand cDNA was synthesized from worm total RNA or *in vitro* synthesized RNA as described (22) followed by purification of the cDNA target, hybridisation of the microarrays overnight at 65°C, washing of the slides and drying of the arrays (22). The slides were scanned 10 using a ScanArray 4000 XL scanner (Perkin-Elmer, USA), and the array data were processed using the GenePix™ Pro 4.0 software package (Axon, USA).

#### D. Assessment of sensitivity and specificity in LNA expression microarrays

To enable direct comparisons between LNA and DNA capture probes in measuring gene 15 expression levels, specific oligonucleotide capture probes for the *Saccharomyces cerevisiae* genes *SWI5* and *THI4* were designed in the 3'-end of the two ORFs. The capture probes were synthesized as 50-mer DNA and corresponding LNA-modified oligonucleotides, respectively, with an LNA substitution at every 3rd nucleotide position. In addition, 40-mer DNA and LNA oligonucleotides were designed as truncated versions of the 50-mer capture probes, along with 20 oligonucleotides with 1 to 5 consecutive mismatches positioned centrally in the 50-mer and 40-mer capture probes. All capture probes were synthesized with an anthraquinone group at the 5'-end and a hexaethyleneglycol dimer linker region (HEG2 spacer arm) enabling photocoupling onto polymer microarray slides as described in US patent No. 6,531,591.

To assess the sensitivity and specificity of the oligonucleotide microarrays, *in vitro* 25 synthesized yeast RNA for either *SWI5* or *THI4* was spiked into *C. elegans* total RNA for cDNA target synthesis followed by hybridization of the microarrays with fluorochrome-labelled cDNA target pools. The incorporation of LNA nucleotides into 50-mer DNA oligonucleotide capture probes results in a 3 to 4-fold increase in fluorescence intensity levels, when hybridized with the spiked, complex cDNA target pools under standard stringency conditions (Figure 45a 30 and c). The sensitivity increase is even more pronounced, 5 to 12-fold, when 40-mer LNA capture probes are employed. None of the yeast capture probes showed cross-hybridization to *C. elegans* cDNA target control without yeast spike RNA under the same conditions.

The specificity of the oligonucleotide capture probes was examined using a panel of LNA mismatch oligonucleotides together with the DNA controls. As demonstrated in Figure 45 35 a and c, the fluorescence intensities obtained with the LNA-modified 40-mer triple mismatch

oligonucleotides show a 3-fold intensity decrease relative to the perfectly matched duplexes. In contrast, the corresponding 40-mer standard DNA capture probes are neither capable of forming stable duplexes nor discriminating between the perfect match and mismatched targets under standard hybridization stringency conditions, resulting in low intensity values from all 5 40-mer DNA capture probes (Figure 45 a and c). Interestingly, mismatch discrimination with the 50-mer LNA probes could be significantly improved by increasing the hybridization temperature from 65 °C to 70 °C (Figure 45 b and d), without compromising their capture sensitivity. By comparison, the signal intensities from all 50-mer DNA capture probes including the perfect match oligonucleotides were reduced under the same conditions (Figure 45 b and d).

10 Considered together, our results strongly support the contention that LNA oligonucleotide capture probes are significantly more sensitive and specific than DNA probes, being able to discriminate between highly homologous (90 %) mRNAs with a 5 to 10-fold increase in sensitivity.

In a typical cell, mRNAs can be subdivided into three kinetic classes: (i) highly 15 abundant (30-90 % of the total mRNA mass, 0.1 % of the sequence complexity); (ii) medium abundant (50 % mass, 2-5 % of complexity); and (iii) low-abundant mRNAs (<1% mass, >90 % of complexity). In addition, alternative splicing has been shown to be prevalent in higher eukaryotes, where at least 50 % of the genes appear to be alternatively spliced, thereby generating additional diversity within the transcriptome. It is thus of utmost importance that the 20 dynamic range, sensitivity and specificity of the expression profiling technology used are optimal, especially when analyzing expression levels of messages and mRNA splice variants belonging to the low-abundant class of high mRNA sequence complexity. A common problem for all DNA oligonucleotide microarrays is the need for an adequate compromise with respect to the sensitivity and specificity of the platform. In the present example LNA oligonucleotide 25 microarrays perform better in expression profiling than microarrays with corresponding DNA probes. Our results clearly demonstrate that both the specificity and sensitivity in target molecule capture can be improved using LNA oligonucleotide microarrays, enabling discrimination between highly homologous mRNAs and alternative splice variants with a simultaneous increase in sensitivity.

30 Figure 45 shows the sensitivity and specificity of LNA oligonucleotide capture probes (black bars) compared to DNA capture probes (white bars) on expression microarrays. Fluorescence intensity is shown in arbitrary units (relative measurements). The arrays comprising 50-mer and 40-mer perfect match and 1-5 mismatch capture probes were hybridized at 65°C in 3xSSC with Cy3-labelled cDNA from 10 µg *C. elegans* total RNA spiked with yeast a) SWI5 RNA and c) 35 THI4 RNA. Demonstration of improved mismatch discrimination with the 50-mer LNA probes

by increasing the hybridization temperature from 65 °C to 70 °C hybridized with Cy3-labelled cDNA from 10 µg *C. elegans* total RNA spiked with yeast b) SWI5 RNA and d) THI4 RNA.

**Example 51. Improved sensitivity in the on-chip capture of yeast HSP78 mRNA using LNA-substituted 25-mer oligonucleotide capture probes.**

**A. Capture probe design**

Unique capture probes for the yeast HSP78 gene were designed using the OligoDesign software, described in Figure 27. The design options used were: (i) length of each oligonucleotide probe was 25 nucleotides; (ii) Blast word length 7; (iii) Blast expectation cut-off 10 off 1000; other options were as default; 24 DNA capture probes were selected. Furthermore, three different LNA-substituted probes were designed based on the sequences of the 24 DNA capture probes, selected by OligoDesign: optimal LNA\_T, optimal LNA\_TC and LNA\_3. In the LNA\_T design the DNA T nucleotides were substituted with LNA T. For the LNA\_TC design, LNA T and C nucleotides were used to substitute DNA t and c. In the LNA\_3 design every third DNA nucleotide was substituted with the corresponding LNA nucleotide. For the LNA\_T and LNA\_TC design, no blocks of LNAs were allowed; in addition the LNAs were substituted in a pattern providing a more narrow duplex melting temperature range compared to the DNA T<sub>m</sub> range. In addition, an equivalent set of capture probes with a single mismatch in the central nucleotide position was designed. Altogether, 192 capture probes were designed including an anthraquinone (AQ29) 5'-modifier and a hexaethyleneglycol dimer (HEG2) at the 5'-end of each probe – as shown in Table 13.

**B. Determination of the duplex melting temperatures (T<sub>m</sub>).**

The duplex melting temperatures of the DNA, LNA T and LNA TC designed probes were measured using the Perkin Elmer Lambda 40 Spectrophotometer and according to Wahlestedt et al. *PNAS* 97/10 2000. All oligos were measured twice and if the replica values deviated more than 1 °C, then a third or a fourth measurement was carried out. The average T<sub>m</sub> for each oligonucleotide duplex is presented in Table 14.

**C. *In vitro* synthesis of fluorochrome-labeled yeast HSP78 RNA**

**C1.** Genomic DNA was prepared from a wild type standard laboratory strain of *Saccharomyces cerevisiae* using the Nucleon MiY DNA extraction kit (Amersham Biosciences) according to supplier's instructions.

**C2.** PCR amplification of the partial yeast gene was done by standard PCR using yeast genomic DNA as template. In the first step of amplification, a forward primer containing a restriction enzyme site and a reverse primer containing a universal linker sequence were used. In this step 20 bp was added to the 3'-end of the amplicon, next to the stop codon. In the second

step of amplification, the reverse primer was exchanged with a nested primer containing a poly-T<sub>20</sub> tail and a restriction enzyme site. The SWI5 amplicon contains 730 bp of the SWI5 ORF plus 20 bp universal linker sequence and a poly-A<sub>20</sub> tail.

The PCR primers used were:

5 YDR258C-For-SacI: acgtgagctcttttgacatgtcagaatttcaag

YDR258C-Rev-Uni: gatccccgggaattgccatgttacttttcagcttctcttcaac

Uni-polyT-BamHI: acgtggatcctttttttttttttttttgatccccgggaattgccatg,

### C3. Plasmid DNA constructs

The PCR amplicon was cut with the restriction enzymes, *EcoRI* + *BamHI*. The DNA fragment 10 was ligated into the pTRIamp18 vector (Ambion) using the Quick Ligation Kit (New England Biolabs) according to the supplier's instructions and transformed into *E. coli* DH-5 $\alpha$  by standard methods.

### C4. DNA sequencing

To verify the cloning of the PCR amplicon, plasmid DNA was sequenced using M13 forward 15 and M13 reverse primers and analysed on an ABI 377.

### C5. Biotin labeling of cRNA

One  $\mu$ g of plasmid containing the HSP78 sequence was linearized with restriction enzyme BamHI (Amersham Pharmacia Biotech, USA) for 2 hours at 37 °C. The RNA was labeled with biotin-CTP and biotin-UTP using the Message AMP aRNA kit from Ambion (USA) according 20 to the manufacturer's instructions. Following hybridisation, the slides were stained with Streptavidin Phycoerythrin (Molecular Probes, S-866, USA) according to the GeneChip Expression Analysis Technical Manual (Affymetrix, USA)

### C6. Fluorochrome labeling of spike RNA

In vitro synthesized spike RNA from the HSP78 plasmid construct was labeled with either Cy3- 25 ULS or Cy5-ULS (Amersham Biosciences, USA) according to the manufacturer's instructions, followed by filtration through a ProbeQuant G50 Micro Column and Microcon30 (Millipore, USA). The labeling efficiency was monitored using the Nanodrop spectrophotometer (Nanodrop Technologies, USA)

### D. Microarray fabrication

30 The microarrays were printed on Immobilizer™ MicroArray Slides (Exiqon, Denmark) using the MicroGrid II from Biorobotics (UK) using a 20  $\mu$ M capture probe solution for each oligonucleotide probe. Four replicas of each capture probe were printed on the slides.

### E. Hybridization with fluorochrome-labelled cRNA

The arrays were hybridized for 16 hours using the following protocol. The labelled RNA samples were hybridized in a hybridization solution (20  $\mu$ L final volume) containing 3 $\times$ SSC (final concentration), 25 mM HEPES, pH 7.0 (final concentration), 1.25  $\mu$ g/ $\mu$ L yeast tRNA, 0.3% SDS. The labeled RNA target sample was filtered in a Millipore 0.22 micron spin column 5 according to the manufacturer's instructions (Millipore, USA), and the probe was denatured by incubating the reaction at 100°C for 2 min. The sample was cooled at 20-25°C for 5 min. by spinning at max speed in a microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was carefully placed on top of the microarray spotted on Immobilizer™ MicroArray Slide and the hybridization mixture was applied to the array from the side. An aliquot of 30  $\mu$ L of 3 $\times$ SSC was 10 added to both ends of the hybridization chamber, and the Immobilizer™ MicroArray Slide was placed in the hybridization chamber. The chamber was sealed watertight and incubated at 45°C, 55°C or 65°C for 16-18 hours submerged in a water bath. After hybridisation, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The Lifterslip coverslip was washed off in 6 $\times$ SSC, pH 7.0 containing 0.1% Tween20 at 50°C 15 for 15 min., followed by washing of the microarrays in 0.4 $\times$ SSC, pH 7.0 at 50°C for 30 min. Finally the slides were washed for 5 seconds in 0.05 $\times$ SSC, pH 7.0. The slides were then dried by centrifugation in a swinging bucket rotor at approximately 200 G for 2 min.

#### F. Data analysis.

Following washing and drying, the slides were scanned using a ScanArray 4000XL scanner 20 (Perkin-Elmer Life Sciences, USA), and the array data were processed using the GenePix™ Pro 4.0 software package (Axon, USA).

#### G. RESULTS

##### G1. Duplex melting temperatures

The  $T_m$  data clearly shows that LNA-substituted oligonucleotide capture probes have a 25 significantly increased average duplex melting temperature compared to the corresponding DNA probes. Furthermore, the difference in melting temperature between the perfectly matched (PM) and single mismatched (MM) probes, designated as  $\Delta T_m$ , is significantly higher than the corresponding  $\Delta T_m$  for DNA probes (Table 12).

Table 12. The average difference in melting temperature between the perfectly matched (PM) and single mismatched (MM) probes in different capture probe designs. The observed difference between the DNA and LNA substituted probes is statistically significant as revealed by a t-Test; Two-Sample Assuming Unequal Variances.

	Average	Max	Min	$\Delta T_m$	St.dev.	T-Test
MM-DNA	60.4	68.5	54.3			
MM-LNA T	67.6	74	58.1			
MM-LNA TC	72.0	79.3	61.6			
PM-DNA	66.3	72.8	61	5.9	1.52	
PM-LNA T	74.2	80.8	66.3	6.6	1.32	0.047
PM-LNA TC	80.0	86.8	71.4	8.0	2.65	0.001 / 0.017

## G2. Microarray hybridization results

Both LNA\_T and LNA\_3 substituted 25-mer probes are capable of providing highly accurate measurements for fold-of-changes in gene expression levels, as depicted in Figure 46. The 10 DNA capture probes did not provide any hybridisation signals under the given microarray hybridisation conditions (Fig. 46). Figure 46 shows the expected (black bars) and observed (white bars) fold-of-change in the expression levels of the Cy3-ULS labelled HSP78 spike RNA as measured by on-chip capture using different oligonucleotide capture probes. In the hybridization experiment, 1 ng HSP78 *in vitro* spike RNA or 200 pg HSP78 *in vitro* spike RNA 15 was used, respectively. Thus, the fold change of the HSP78 RNA in the two hybridizations in the comparison is 5-fold. Fourteen additional synthetic *in vitro* mRNA spike controls were included in the hybridisation solution as a semi-complex background RNA mixture. Seven of these spikes were used as normalization controls, the other seven were used as negative controls. Hybridization temperature was 65°C for 16 hours, and post-hybridization washes as 20 described above. Under these conditions the DNA capture probes didnot produce hybridization signals. Figure 47 shows measured intensity levels by on-chip capture using three different 25-mer oligonucleotide capture probe designs. One (1) ng biotin-labeled HSP78 target was used in the hybridization experiments, followed by staining with Streptavidin Phycoerythrin. The LNA\_T and LNA\_TC substituted 25-mer capture probes show a significantly enhanced on-chip 25 capture of the HSP78 RNA target, compared to the DNA 25-mer control probes under four different hybridization stringency conditions.



Table 13. Design of the yeast HSP78 capture probes. YDR258C denotes the ORF name of the *S. cerevisiae* HSP78 gene. The numbers refer to the nucleotide position from the 3'-end of the HSP78 mRNA sequence. PM = perfectly matched probe, MM = single mismatch probe, LNA substitutions are depicted by capital letters, <sup>m</sup>C denotes LNA methyl-C

Oligo name	Sequence	Oligo name	Sequence
YDR258C_PM_043	tttgtagcacgacaagcttagtat	YDR258C_PM_043T	TTTggTagcacgacaagcTTagTat
YDR258C_PM_078	cactctaacagtttcgccgtticta	YDR258C_PM_078T	cacTcTaacagtTtcgccgTTTcTa
YDR258C_PM_124	gttgccatggagttcaaaatctgtc	YDR258C_PM_124T	gTTgccatGgagTtcaaaaTcTgTc
YDR258C_PM_164	tcaatggccttgaccataaattg	YDR258C_PM_164T	TcaaTggccTTgcaccaTaTaaTTg
YDR258C_PM_201	tcagttagccaatcccttcgcttcat	YDR258C_PM_201T	TcagTTagccaaTccTTgcgTTcat
YDR258C_PM_249	tttttcggccaacgatcttgaatt	YDR258C_PM_249T	TtTTTcgccaacgaTcTTgaaTt
YDR258C_PM_295	attgacctcaaaactttcttgata	YDR258C_PM_295T	aTTgaccTcaaaacTTTcTTggaTa
YDR258C_PM_356	gatgaactcaggtggataggatctt	YDR258C_PM_356T	gaTgaacTcaggTggaTaggaTcTt
YDR258C_PM_424	accatcatcacccaacttgtgtcg	YDR258C_PM_424T	accaTcaTcacccaacTTTgTgTcg
YDR258C_PM_433	cccaactttgtgtcgtttaataaaa	YDR258C_PM_433T	cccaacTTTgTgTcgTTTaaTaaaa
YDR258C_PM_486	caatgatcgtgttacggaatcaac	YDR258C_PM_486T	caaTgaTcgTgtTacggaatTcaac
YDR258C_PM_515	tgccctagggaatcggcagcttac	YDR258C_PM_515T	TggccTagggaaTcggTcagcTTac
YDR258C_PM_566	ttggaacaatcggggtgcgctttt	YDR258C_PM_566T	TTggaacaTcggggTgcgcTTTTt
YDR258C_PM_569	gaaacatcggggtgcgcttttcaa	YDR258C_PM_569T	gaaacaTcggggTgcgcTTTTTcaa
YDR258C_PM_604	aaaacgacagcataaggctttctt	YDR258C_PM_604T	aaaacgacagcaTaaggcTTTcTTc
YDR258C_PM_631	gacagcctcagtttaattggccacca	YDR258C_PM_631T	gacagccTcagtTaaTTggccacca
YDR258C_PM_686	ccgalttaaacgagagacagatgct	YDR258C_PM_686T	ccgaTTaaacgagagacagTaTgct
YDR258C_PM_757	atcanatagggaattcagctaaagcc	YDR258C_PM_757T	aTcaaaTaggaaTTcagcTaaagcc
YDR258C_PM_813	gacctaaagaacataaagctggcaat	YDR258C_PM_813T	gaccTaagaacaTaaagcTggcaat
YDR258C_PM_823	ataaagctggcaataggctctttt	YDR258C_PM_823T	aTaaagcTggcaaTaggTcTcTTt
YDR258C_PM_870	agacgtacagcatcagaatagcag	YDR258C_PM_870T	agacgTacagcaTcagaaaTagcag
YDR258C_PM_888	aaatagcagcaatggcctcgtcttg	YDR258C_PM_888T	aaaTagcagcaaTggccTcgTcTTg
YDR258C_PM_890	atagcagcaatggcctcgtcttggc	YDR258C_PM_890T	aTagcagcaaTggccTcgTcTTggc
YDR258C_PM_896	gcaatggcctcgtcttgccaacga	YDR258C_PM_896T	gcaaTggccTcgTcTTggccaacga
YDR258C_PM_043	TtTggTagmCamCgamCaagmCrT	YDR258C_PM_LNA3_043	TtTggtAgcAcgAcaAgcTtaGtat
TC	agTat		
YDR258C_PM_078	mCamCtmCtaamCagtTtcmCcgT	YDR258C_PM_LNA3_078	mCacTctAacAgtTtcGccGttTcta
TC	tTmCTa		
YDR258C_PM_124	gTtgmCmCaTggagTtmCaaaaTm	YDR258C_PM_LNA3_124	GttGccAtgGagTtcAaaAtcTgtc
TC	CtgTc		
YDR258C_PM_164	TmCaaTggmCcTTgmCacmCaTa	YDR258C_PM_LNA3_164	TcaAtgGccTtgmCacmCatAtaAttg
TC	TaaTTg		
YDR258C_PM_201	TmCagTtagcmCaaTcmCtTmCgm	YDR258C_PM_LNA3_424	AccAtcAtcAccmCaamCttTgtGtcg
TC	CttmCat		
YDR258C_PM_249	TtTtTmCggmCmCaaamCgatmCt	YDR258C_PM_LNA3_486	mCaaTgaTcgTgtTacGgaAatmCaac
TC	TgaaTt		

YDR258C_PM_295	aTTgamCmCTmCaaaamCTTtmC	YDR258C_PM_LNA3_515	TggmCctAggGaaTcgGtcAgcTlac
TC	TTggaTa	YDR258C_PM_LNA3_566	TtgGaaAcaTcgGggTgcGctTitt
YDR258C_PM_356	gaTgaamCTmCaggTggaTaggaT	YDR258C_PM_LNA3_569	GaaAcaTcgGggTgcGctTttTcaa
TC	mCTt	YDR258C_PM_LNA3_604	AaaAcgAcaGcaTaaGgcTttmCttc
YDR258C_PM_424	amCcaTcaTcacmCmCaaccTtgTgt	YDR258C_PM_LNA3_757	AtcAaaTagGaaTtcAgcTaaAgcc
TC	mCg	YDR258C_PM_LNA3_813	GacmCtaAgaAcaTaaAgcTggmCaat
YDR258C_PM_433	mCmCmCaacTThgTgTcgTTTaaT	YDR258C_PM_LNA3_823	AtaAagmCtgGcaAtaGgtmCtcTitt
TC	aaaa	YDR258C_PM_LNA3_870	AgamCgtAcaGcaTcaGaaAtaGcag AaaTagmCagmCaaTggmCctmCgtmCt
YDR258C_PM_486	mCaaTgaTmCgTgtTamCggaaaT	YDR258C_PM_LNA3_888	tg
TC	mCaac	YDR258C_PM_LNA3_890	AtaGcaGcaAtgGccTcgTctTggc
YDR258C_PM_515	TggmCcTagggaaTmCggTcagmCt	YDR258C_PM_LNA3_896	GcaAtgGccTcgTctTggmCcaAcga
TC	Tac	YDR258C_PM_LNA3_631	GacAgcmCtcAgTaaTtgGccAcca
YDR258C_PM_566	TTggaaamCatmCggggTgmCgctT	YDR258C_PM_LNA3_686	mCcgAttAaamCgaGagAcaGtaTgct
TC	tTt	YDR258C_PM_LNA3_356	GatGaamCtcAggTggAtaGgaTctt
YDR258C_PM_569	gaaamCaTmCggggTgmCgctttTtm	YDR258C_PM_LNA3_201	TcaGttAgcmCaaTccTtcGctTcat
TC	Caa	YDR258C_PM_LNA3_249	TttTtcGgcmCaaAcgAtcTtgAatt
YDR258C_PM_604	aaaamCgamCagmCaTaaggmCTt	YDR258C_PM_LNA3_295	AttGacmCtcAaaActTtcTtgGata
TC	TmCtTc	YDR258C_PM_LNA3_433	mCccAacTttGtgTcgTttAatAaaa
YDR258C_PM_631	gamCagmCcTcagtTaaTTggcmCa	YDR258C_MM_043T	TTTggTagcacgtcaagcTTagTat
TC	cmCa	YDR258C_MM_078T	cacTcTaacagatcgcgcTTTcTa
YDR258C_PM_686	mCmCgaTTaaamCgagagamCagT	YDR258C_MM_124T	gTTgccaTggagatcaaaaTcTgTc
TC	aTgmCt	YDR258C_MM_164T	TcaaTggccTTggaccaTaTaaTTg
YDR258C_PM_757	aTmCaaaTaggaaTtmCagmCTaaa	YDR258C_MM_201T	TcagTTagccaaccTTcgcTTcal
TC	gmCc	YDR258C_MM_249T	TtTTTcgccaatcgaTcTTgaaTt
YDR258C_PM_813	gamCmCTaagaamCaTaaagmCTg	YDR258C_MM_295T	aTTgaccTcaaatcTTTcTTggaTa
TC	gmCaat		
YDR258C_PM_823	aTaaagmCTggmCaaTaggTmCTc		
TC	TTTt		
YDR258C_PM_870	agamCgTamCagmCaTmCagaaaT		
TC	agmCag		
YDR258C_PM_888	aaaTagmCagmCaaTggmCcTmCg		
TC	TctTg		
YDR258C_PM_890	aTagmCagcaaTggmCcTcgtmCtT		
TC	ggc		
YDR258C_PM_896			
TC	gcaatggcctmCgTmCttggccaacga		
YDR258C_MM_043	tttgtagcacgtcaagcctagtat		
YDR258C_MM_078	cactctaacagtatcgcgtttcta		
YDR258C_MM_124	gttgccatggagatcaaaatctgic		
YDR258C_MM_164	tcaatggccittggaccalataattg		
YDR258C_MM_201	tcagttagccaacettcgttcat		
YDR258C_MM_249	ttttcgccaatcgatctgaatl		
YDR258C_MM_295	attgacctcaaatctttctggata		

YDR258C_MM_356	gatgaacTcaggaggataggatctt	YDR258C_MM_356T	gaTgaacTcaggaggataggatctt
YDR258C_MM_424	accatcatcaccgaactttgtgtcg	YDR258C_MM_424T	accaTcaTcaccgaacTTTgTgTcg
YDR258C_MM_433	cccaactttgtgacgtttaataaaa	YDR258C_MM_433T	cccaacTTTgTgacgTTTaaTaaaa
YDR258C_MM_486	caatgatcgtgtaacggaaatcaac	YDR258C_MM_486T	caaTgaTcgTgtaacggaaaTcaac
YDR258C_MM_515	tggcctagggaacggcagcttac	YDR258C_MM_515T	TggccTagggaaacggTcagcTTac
YDR258C_MM_566	tggaaacatcgcggtgcgttttt	YDR258C_MM_566T	TTggaaacaTcgcggTgcgcTTTTt
YDR258C_MM_569	gaaacatcggggagcgtttttcaa	YDR258C_MM_569T	gaaacaTcggggagcgtTTTTcaa
YDR258C_MM_604	aaaacgacagcaaaaggctttcttc	YDR258C_MM_604T	aaaacgacagcaaaaggcTTTcTTc
YDR258C_MM_631	gacagcctcagtaaatggccacca	YDR258C_MM_631T	gacagccTcagtaaaTTggccacca
YDR258C_MM_686	ccgattaaacgacagacagtatgct	YDR258C_MM_686T	ccgaTTaaacgacagacagTaTgct
YDR258C_MM_757	atcaaatagggaatcagctaaagcc	YDR258C_MM_757T	aTcaaaTagggaatcagcTaaagcc
YDR258C_MM_813	gacctagaacaaaaagctggcaat	YDR258C_MM_813T	gaccTaagaacaaaaagcTggcaat
YDR258C_MM_823	ataaagctggcattaggtctctttt	YDR258C_MM_823T	aTaaagcTggcaaaaggTcTcTTTt
YDR258C_MM_870	agacgtacagcaacagaaatagcag	YDR258C_MM_870T	agacgTacagcaacagaaaTagcag
YDR258C_MM_888	aaatagcagcaaaaggcctcgtcttg	YDR258C_MM_888T	aaaTagcagcaaaaggccTcgTcTTg
YDR258C_MM_890	atagcagcaatgccctcgtcttgcc	YDR258C_MM_890T	aTagcagcaaTgcccTcgTcTTggc
YDR258C_MM_896	gcaatggcctcgacttggccaacga	YDR258C_MM_896T	gcaaTggccTcgactTggccaacga
YDR258C_MM_043	TtTggTagmCamCgtcaagmCtTag	YDR258C_MM_LNA3_043	TttGgtAgcAcgTcaAgcTtaGtat
TC	Tat		
YDR258C_MM_078	mCamCtmCtaamCagtatcgmCcgT	YDR258C_MM_LNA3_078	mCacTctAacAgtAtcGccGttTcta
TC	tTmCTa		
YDR258C_MM_124	gTlgmCmCaTggagatmCaaaaTmC	YDR258C_MM_LNA3_124	GttGccAtgGagAtcAaaAtcTgtc
TC	tgTc		
YDR258C_MM_164	TmCaaTggmCcTTggacmCaTaTa	YDR258C_MM_LNA3_164	TcaAtgGccTtgGacmCatAtaAttg
TC	aTTg		
YDR258C_MM_201	lmCagTlagcmCaaacctTcgmCttm	YDR258C_MM_LNA3_201	TcaGttAgcmCaaAccTtcGctTcat
TC	Cat		
YDR258C_MM_249	TtTtTmCggmCmCaatcgatmCtTg	YDR258C_MM_LNA3_249	TttTtcGgcmCaaTcgAtcTtgAatt
TC	aaTt		
YDR258C_MM_295	aTTgamCmCTmCaaatCTTmCTT	YDR258C_MM_LNA3_295	AttGacmCtcAaaTctTtcTtgGata
TC	ggaTa		
YDR258C_MM_356	gaTgaamCTmCaggaggaTaggaTm	YDR258C_MM_LNA3_356	GatGaamCtcAggAggAiaGgaTctt
TC	CTt		
YDR258C_MM_424		YDR258C_MM_LNA3_424	AccAtcAtcAccGaamCttTgtGtcg
TC	amCcaTcaTcaccacaccTtgTgtmCg		
YDR258C_MM_433	mCmCmCaacTTtgTgacgTTTaaT	YDR258C_MM_LNA3_433	mCccAacTttGtgAcgTttAatAaaa
TC	aaaa		
YDR258C_MM_486	mCaaTgaTmCgTgttamCggaaaTm	YDR258C_MM_LNA3_486	mCaaTgaTcgTgtAacGgaAatmCaac
TC	Caac		
YDR258C_MM_515	TggmCcTagggaaacggTcagmCtTa	YDR258C_MM_LNA3_515	TggmCctAggGaaAcgGtcAgcTtac
TC	c		
YDR258C_MM_566	TTggaaamCatmCgcggTgmCgctTt	YDR258C_MM_LNA3_566	TtgGaaAcaTcgmCggTgcGctTttt
TC	Tt		

YDR258C_MM_569	gaaamCaTmCggggagmCgcttTtm	YDR258C_MM_LNA3_569	GaaAcaTcgGggAgcGctTtTcaa
TC	Caa		
YDR258C_MM_604	aaaamCgamCagmCaaaaggmCTtT	YDR258C_MM_LNA3_604	AaaAcgAcaGcaAaaGgcTtmCttc
TC	mCtTc		
YDR258C_MM_631	gamCagmCcTcagtaaaTTggcmCa	YDR258C_MM_LNA3_631	GacAgcmCtcAgtAaaTtgGccAcca
TC	cmCa		
YDR258C_MM_686	mCmCgaTTaaamCgacagamCagT	YDR258C_MM_LNA3_686	mCcgAttAaamCgamCagAcaGtaTgct
TC	aTgmCt		
YDR258C_MM_757	aTmCaaaTaggaaatmCagmCTaaag	YDR258C_MM_LNA3_757	AtcAaaTagGaaAtcAgcTaaAgcc
TC	mCc		
YDR258C_MM_813	gamCmCTaagaamCaaaaagmCTg	YDR258C_MM_LNA3_813	GacmCtaAgaAcaAaaAgcTggmCaat
TC	gmCaat		
YDR258C_MM_823	aTaaagmCTggmCattaggTmCTcT	YDR258C_MM_LNA3_823	AtaAagmCtgGcaTtaGgtmCtcTttt
TC	TTt		
YDR258C_MM_870	agamCgTamCagmCaacagaaaTag	YDR258C_MM_LNA3_870	AgamCgtAcaGcaAcaGaaAtaGcag
TC	mCag		
YDR258C_MM_888	aaaTagmCagmCaaaggmCcTmCg		AaaTagmCagmCaaAggmCctmCgtmCt
TC	TctTg	YDR258C_MM_LNA3_888	tg
YDR258C_MM_890			
TC	aTagmCagcaaTgcccTcgtmCrTggc	YDR258C_MM_LNA3_890	AtaGcaGcaAtgmCccTcgTctTggc
YDR258C_MM_896	gmCaaTggcctmCgactggccaamCg		
TC	a	YDR258C_MM_LNA3_896	GcaAtgGccTcgActTggmCcaAcga

Table 14. Duplex melting temperatures (T<sub>m</sub>) for the 144 different 25-mer oligonucleotide capture probes. The design column denotes the sequence design of the probe. PM = perfectly matched probe, MM = single mismatch probe, LNA substitutions are depicted by capital letters, 5 <sup>m</sup>C denotes LNA methyl-C

Probe number	Oligonucleotide target name	Complementary target sequence	Design	Average duplex melting temp. (°C)
12696	YDR258C_Tm_predic_043	tttgtagcacgtcaagcttagtat	MM-DNA	59.2
12699	YDR258C_Tm_predic_078	cactctaacagtatcgccgtttcta	MM-DNA	61.1
12694	YDR258C_Tm_predic_124	gttgccatggagatcaaaatctgtc	MM-DNA	60
12700	YDR258C_Tm_predic_164	tcaatggccttggaccatataattg	MM-DNA	59.5
12693	YDR258C_Tm_predic_201	tcagttagccaaccttcgcttcat	MM-DNA	61.4
12698	YDR258C_Tm_predic_249	ttttcggccaatcgatcttgaatt	MM-DNA	58.8
12702	YDR258C_Tm_predic_295	attgacctcaaatctttcttgata	MM-DNA	54.5
12692	YDR258C_Tm_predic_356	gatgaactcaggagataggtatctt	MM-DNA	58.2
12680	YDR258C_Tm_predic_424	accatcatcaccgaactttgtgtcg	MM-DNA	63.6
12703	YDR258C_Tm_predic_433	cccaactttgtgacgtttaataaaa	MM-DNA	56.2
12681	YDR258C_Tm_predic_486	caatgatcgtgtaacggaatcaac	MM-DNA	59.3

12682	YDR258C_Tm_predic_515	tgccctagggaacgggcagcttac	MM-DNA	65.2
12683	YDR258C_Tm_predic_566	tggaaacatcgcggtgcgtttt	MM-DNA	61.5
12684	YDR258C_Tm_predic_569	gaaacatcggggagcgcttttcaa	MM-DNA	63.8
12701	YDR258C_Tm_predic_604	aaaacgacagcaaaaggcttcttc	MM-DNA	58.8
12685	YDR258C_Tm_predic_631	gacagcctcagtaattggccacca	MM-DNA	64.8
12686	YDR258C_Tm_predic_686	ccgattaacgacagacagatgct	MM-DNA	57.1
12687	YDR258C_Tm_predic_757	atcaaataggaaatcagctaaagcc	MM-DNA	54.3
12688	YDR258C_Tm_predic_813	gacctaagaacaaaagctggcaat	MM-DNA	58.2
12697	YDR258C_Tm_predic_823	ataaagctggcattaggctctttt	MM-DNA	58.4
12695	YDR258C_Tm_predic_870	agacgtacagcaacagaaatgcag	MM-DNA	61.2
12689	YDR258C_Tm_predic_888	aaatagcagcaaaagcctcgcttg	MM-DNA	64
12690	YDR258C_Tm_predic_890	atagcagcaatgccctcgcttgcc	MM-DNA	62.9
12691	YDR258C_Tm_predic_896	gcaatggcctcgacttggccaacga	MM-DNA	68.5
12720	YDR258C_Tm_predic_043T	TTTggTagcagctcaagcTTTagTat	MM-T	68.6
12723	YDR258C_Tm_predic_078T	cacTcTaacagtatcgcgTTTcTa	MM-T	69.7
12718	YDR258C_Tm_predic_124T	gTTgccatggagatcaaaaTcTgTc	MM-T	69.2
12724	YDR258C_Tm_predic_164T	TcaaTggccTTggaccaTaTaaTTg	MM-T	69.7
12717	YDR258C_Tm_predic_201T	TcagTTagccaaccTTcgCTTcat	MM-T	69.4
12722	YDR258C_Tm_predic_249T	TtTTTcgccaatcgaTcTTgaaTt	MM-T	65.9
12726	YDR258C_Tm_predic_295T	aTTgaccTcaaatCTTcTTggaTa	MM-T	65.1
12716	YDR258C_Tm_predic_356T	gaTgaacTcaggaggaTaggaTcTt	MM-T	64.9
12704	YDR258C_Tm_predic_424T	accaTcaTcaccgaacTTTgTgTcg	MM-T	74
12727	YDR258C_Tm_predic_433T	cccaacTTTgTgacgTTTaaTaaaa	MM-T	66.5
12705	YDR258C_Tm_predic_486T	caaTgaTcgTgtaacgaaaTcaac	MM-T	65.2
12706	YDR258C_Tm_predic_515T	TggccTagggaacggTcagcTTac	MM-T	71.6
12707	YDR258C_Tm_predic_566T	TTggaaacaTcgcgTgcgcTTTTt	MM-T	68.6
12708	YDR258C_Tm_predic_569T	gaaacaTcggggagcgcTTTTTcaa	MM-T	69.9
12725	YDR258C_Tm_predic_604T	aaaacgacagcaaaaggcTTTcTTc	MM-T	65.9
12709	YDR258C_Tm_predic_631T	gacagccTcagtaaaTTggccacca	MM-T	68.8
12710	YDR258C_Tm_predic_686T	ccgaTTaaacgacagacagTaTgct	MM-T	63.5
12711	YDR258C_Tm_predic_757T	aTcaaaTaggaaatcagcTaaagcc	MM-T	58.1
12712	YDR258C_Tm_predic_813T	gaccTaagaacaaaaagcTggcaat	MM-T	61.1
12721	YDR258C_Tm_predic_823T	aTaaagcTggcaaaaggTcTcTTTTt	MM-T	13+14 67.2
12719	YDR258C_Tm_predic_870T	agacgTacagcaacagaaaTagcag	MM-T	65
12713	YDR258C_Tm_predic_888T	aaaTagcagcaaaaggccTcgTcTTg	MM-T	70.4
12714	YDR258C_Tm_predic_890T	aTagcagcaaTgccTcgTcTTggc	MM-T	71.3
12715	YDR258C_Tm_predic_896T	gcaaTggccTcgacTTggccaacga	MM-T	73.7
12744	YDR258C_Tm_predic_043TC	TtTggTagmCamCgtaagmCrTagTat	MM-TC	73.3
12747	YDR258C_Tm_predic_078TC	mCamCtmCtaamCagtatcgmcgTtTmCTa	MM-TC	75.2
12742	YDR258C_Tm_predic_124TC	gTtgmCmCaTggagatmCaaaaTmCtgTc	MM-TC	61.6
12748	YDR258C_Tm_predic_164TC	TmCaaTggmCcTTggacmCaTaTaaTTg	MM-TC	74.8
12741	YDR258C_Tm_predic_201TC	tmCagTtagcmCaaacctTcgmcTtmCat	MM-TC	70.6
12746	YDR258C_Tm_predic_249TC	TtTtTmCggmCmCaatcgatmCrTgaaTt	MM-TC	71
12750	YDR258C_Tm_predic_295TC	aTTgamCmCtmCaaatCTTmCTTggaTa	MM-TC	72.2

12740	YDR258C_Tm_predic_356TC	gaTgaamCTmCaggaggaTaggaTmCTt	MM-TC	70.4
12728	YDR258C_Tm_predic_424TC	amCcaTcaTcacccaaccTtgTgtmCg	MM-TC 13+16	70.2
12751	YDR258C_Tm_predic_433TC	mCmCmCaacTTtgTgacgTTTaaTaaaa	MM-TC	67.6
12730	YDR258C_Tm_predic_515TC	TggmCcTagggaaacggTcagmCtTac	MM-TC	75.5
12731	YDR258C_Tm_predic_566TC	TTggaaamCatmCgchgTgmCgctTtTt	MM-TC	72
12732	YDR258C_Tm_predic_569TC	gaaamCaTmCggggagmCgcttTmCaa	MM-TC	74.8
12749	YDR258C_Tm_predic_604TC	aaaaCgamCagmCaaaaggmCTtTmCtTc	MM-TC	72
12733	YDR258C_Tm_predic_631TC	gamCagmCcTcagtaaaTTggcmCacmCa	MM-TC	77.4
12734	YDR258C_Tm_predic_686TC	mCmCgaTTaaamCgacagamCagTaTgmCt	MM-TC	70.2
12735	YDR258C_Tm_predic_757TC	aTmCaaaTaggaaatmCagmCTaaagmCc	MM-TC	64.6
12736	YDR258C_Tm_predic_813TC	gamCmCTaagaamCaaaaagmCTggmCaat	MM-TC	71.4
12745	YDR258C_Tm_predic_823TC	aTaaagmCTggmCattaggTmCTcTTTt	MM-TC	74
12743	YDR258C_Tm_predic_870TC	agamCgTamCagmCaacagaaaTagmCag	MM-TC	73.6
12737	YDR258C_Tm_predic_888TC	aaaTagmCagmCaaaggmCcTmCgTctTg	MM-TC	79.3
12738	YDR258C_Tm_predic_890TC	aTagmCagcaaTgcccTcgtmCtTggc	MM-TC	71.5
12739	YDR258C_Tm_predic_896TC	gcaatggcctmCgamCttggccaacga	MM-TC	73.1
12768	YDR258C_Tm predi_043_PM	tttgtagcacgacaagcttagtat	PM-DNA	65.7
12771	YDR258C_Tm predi_078_PM	cactctaacagttcgcctticta	PM-DNA	66.3
12766	YDR258C_Tm predi_124_PM	gttgccatggagttcaaaatctgtc	PM-DNA	65.8
12772	YDR258C_Tm predi_164_PM	tcaatggccttgccatataattg	PM-DNA	64
12765	YDR258C_Tm predi_201_PM	tcagtagccaatcttcgcttcat	PM-DNA	66.1
12770	YDR258C_Tm predi_249_PM	ttttcggccaacagatcttgaatt	PM-DNA	65
12774	YDR258C_Tm predi_295_PM	attgacctcaaaaactttcttgata	PM-DNA	61.8
12764	YDR258C_Tm predi_356_PM	gaigaactcaggtaggtagatcct	PM-DNA	64
12752	YDR258C_Tm predi_424_PM	accatcatcacccaactttgtgtcg	PM-DNA	67.5
12775	YDR258C_Tm predi_433_PM	cccaactttgtgtcttlaataaaa	PM-DNA	61
12753	YDR258C_Tm predi_486_PM	caatgatcgtgtacggaatcaac	PM-DNA	64.2
12754	YDR258C_Tm predi_515_PM	tgccctagggaatcggtagcttac	PM-DNA	70.1
12755	YDR258C_Tm predi_566_PM	ttggaacatcgggtgcgtctttt	PM-DNA	70.8
12756	YDR258C_Tm predi_569_PM	gaaacatcgggtgcgtcttttcaa	PM-DNA	68.7
12773	YDR258C_Tm predi_604_PM	aaaacgacagcataaggctttcttc	PM-DNA	64.5
12757	YDR258C_Tm predi_631_PM	gacagctcagtaattggccacca	PM-DNA	70.2
12758	YDR258C_Tm predi_686_PM	ccgattaacgagagacagatgct	PM-DNA	65.4
12759	YDR258C_Tm predi_757_PM	atcaaataggaattcagctaaagcc	PM-DNA	61.5
12760	YDR258C_Tm predi_813_PM	gacctagaacataaagctggcaat	PM-DNA	63.5
12769	YDR258C_Tm predi_823_PM	ataaagctggcaataggctctttt	PM-DNA	63.6
12767	YDR258C_Tm predi_870_PM	agacgtacagcatcagaatagcag	PM-DNA	66.6
12761	YDR258C_Tm predi_888_PM	aaatagcagcaatggcctcgtcttg	PM-DNA	69.2
12762	YDR258C_Tm predi_890_PM	atagcagcaatggcctcgtcttgcc	PM-DNA	72.7
12763	YDR258C_Tm predi_896_PM	gcaatggcctcgtcttgccaacga	PM-DNA	72.8
12792	YDR258C_Tm predi_043T_PM	TTTggTagcacgacaagcTTagTat	PM-T	74.2
12795	YDR258C_Tm predi_078T_PM	cacTcTaacagtTtcgccgTTTcTa	PM-T	75.5
12790	YDR258C_Tm predi_124T_PM	gTTgccatggagTtcaaaaTcTgTc	PM-T	74.7
12796	YDR258C_Tm predi_164T_PM	TcaaTggccTTgcaccaTaTaaTTg	PM-T	74

12789	YDR258C_Tm predic_201T_PM	TcagTTagccaaTccTTcgcTTcat	PM-T	75.7
12794	YDR258C_Tm predic_249T_PM	TtTTTCggccaacgaTcTTgaaTt	PM-T	71
12798	YDR258C_Tm predic_295T_PM	aTTgaccTcaaacTTTcTTggaTa	PM-T	70.2
12788	YDR258C_Tm predic_356T_PM	gaTgaacTcaggTggaTaggaTcTt	PM-T	71
12776	YDR258C_Tm predic_424T_PM	accaTcaTcacccaacTTTgTgTcg	PM-T	79.2
12799	YDR258C_Tm predic_433T_PM	cccaacTTTgTgTcgTTTaaTaaaa	PM-T	75.1
12777	YDR258C_Tm predic_486T_PM	caaTgaTcgTgtTaccgaaaTcaac	PM-T	72.2
12778	YDR258C_Tm predic_515T_PM	TggccTagggaaTcggTcagcTTac	PM-T	76.9
12779	YDR258C_Tm predic_566T_PM	TTggaacaTcggggTgcgcTTTTt	PM-T	76.7
12780	YDR258C_Tm predic_569T_PM	gaaacaTcggggTgcgcTTTTTcaa	PM-T	78
12797	YDR258C_Tm predic_604T_PM	aaaacgacagcaTaaggcTTTcTTc	PM-T	71.7
12781	YDR258C_Tm predic_631T_PM	gacagccTcagtTaaTTggccacca	PM-T	75.4
12782	YDR258C_Tm predic_686T_PM	ccgaTTaaacgagagacagTaTgct	PM-T	72.2
12783	YDR258C_Tm predic_757T_PM	aTcaaaTaggaaTTcagcTaaagcc	PM-T	66.3
12784	YDR258C_Tm predic_813T_PM	gaccTaagaacaTaaagcTggcaat	PM-T	67.5
12793	YDR258C_Tm predic_823T_PM	aTaaagcTggcaaTaggTcTcTTTTt	PM-T	74.3
12791	YDR258C_Tm predic_870T_PM	agacgTcacgcaTcagaaaTagcag	PM-T	71.4
12785	YDR258C_Tm predic_888T_PM	aaaTagcagcaaTggccTcgTcTTg	PM-T	77.2
12786	YDR258C_Tm predic_890T_PM	aTagcagcaaTggccTcgTcTTggc	PM-T	80.1
12787	YDR258C_Tm predic_896T_PM	gcaaTggccTcgTcTTggccaacga	PM-T	80.8
12816	YDR258C_Tm predic_043TC_PM	TtTggTagmCamCgamCaagmCtTagTat	PM-TC	79
12819	YDR258C_Tm predic_078TC_PM	mCamCtmCtaamCagtTicgmCcgTtTmCTa	PM-TC	81.9
12814	YDR258C_Tm predic_124TC_PM	gTigmCmCaTggagTmCaaaaTmCtgTc	PM-TC	78.3
12820	YDR258C_Tm predic_164TC_PM	TmCaaTggmCcTTgmCacmCaTaTaaTTg	PM-TC	83.5
12813	YDR258C_Tm predic_201TC_PM	TmCagTtagcmCaaTcmCtTmCgmCttmCat	PM-TC	81.8
12818	YDR258C_Tm predic_249TC_PM	TtTtTmCggmCmCaaamCgatmCtTgaaTt	PM-TC	77.8
12822	YDR258C_Tm predic_295TC_PM	aTTgamCmCTmCaaamCTTtmCTTggaTa	PM-TC	75.6
12812	YDR258C_Tm predic_356TC_PM	gaTgaamCtmCaggTggaTaggaTmCTt	PM-TC	77.3
12800	YDR258C_Tm predic_424TC_PM	amCcaTcaTcacmCmCaaccTtgTgtmCg	PM-TC	74.2
12823	YDR258C_Tm predic_433TC_PM	mCmCmCaacTTtgTgTcgTTTaaTaaaa	PM-TC	74.8
12729	YDR258C_Tm predic_486TC_PM	mCaaTgaTmCgTgtamCggaamTmCaac	PM-TC	75.5
12802	YDR258C_Tm predic_515TC_PM	TggmCcTagggaaTmCggTcagmCtTac	PM-TC	83.6
12803	YDR258C_Tm predic_566TC_PM	TTggaaamCamCggggTgmCgctTtTt	PM-TC	80.8
12804	YDR258C_Tm predic_569TC_PM	gaaamCaTmCggggTgmCgcttTtTmCaa	PM-TC	83.2
12821	YDR258C_Tm predic_604TC_PM	aaaamCgamCagmCaTaaggmCTtTmCtTc	PM-TC	79
12805	YDR258C_Tm predic_631TC_PM	gamCagmCcTcagtTaaTTggcmCacmCa	PM-TC	82.5
12806	YDR258C_Tm predic_686TC_PM	mCmCgaTTaaamCgagagamCagTaTgmCt	PM-TC	79.4
12807	YDR258C_Tm predic_757TC_PM	aTmCaaaTaggaaTmCagmCTaaagmCc	PM-TC	71.4
12808	YDR258C_Tm predic_813TC_PM	gamCmCTaagaamCaTaaagmCTggmCaat	PM-TC	78.9
12817	YDR258C_Tm predic_823TC_PM	aTaaagmCTggmCaaTaggTmCTcTTTTt	PM-TC	81.2
12815	YDR258C_Tm predic_870TC_PM	agamCgTmCagmCaTmCagaaaTagmCag	PM-TC	81.9
12809	YDR258C_Tm predic_888TC_PM	aaaTagmCagmCaaTggmCcTmCgTctTg	PM-TC	86.8
12810	YDR258C_Tm predic_890TC_PM	aTagmCagcaaTggmCcTcgtmCtTggc	PM-TC	83
12811	YDR258C_Tm predic_896TC_PM	gcaatggcctmCgTmCtggccaacga	PM-TC	79.3

**Example 52. Performance analysis of LNA substituted oligonucleotide capture probes designed to detect splice variants in complex RNA pools.**

A. Oligonucleotide design for microarrays. The methods for designing exon-specific internal 5 oligonucleotide capture probes has been described in example 2.

**A1. Design of the LNA-modified capture probes**

For the internal LNA-modified oligonucleotide capture probes, every third DNA nucleotide was substituted with an LNA nucleotide. The probes designed to capture the splice junction of the recombinant splice variants were designed with LNA substitutions at every third 10 nucleotide position. All capture probes are shown in Table 15.

Table 15. Internal, exon-specific and merged, exon-exon splice junction specific oligonucleotide capture probes used in the example 51. Capital letters denote LNA nucleotides and <sup>m</sup>C LNA methyl-cytosine

15

EQ No	Oligo Name	Sequence
10716	>gene78.01a	cctgaaagtagatttgatttccgaaacgccttctcccgttctaagtc
10717	>gene78.01b	catataccacaaatagtcctcaaaaatcacaagaaaactcacaacactg
10718	>gene78.03a	gatttgacgagggtgtaaaaagtagaaaacgtggaattaaaagggtctc
10719	>gene78.03b	ccaatgaaaactaatcaagggttaaacgtggatcccatggcaattcccggg
10720	>gene78.m0103	cacaacactgcccagaggttcaatcgataaatatgtgaaggaaatgcctg
10721	>gene78.m01INS3	caacactgcccagaggttcaatcgatccgatgatcctaataagggcgccc
10722	>gene78.mINS303	gtccagatcgtccatcatagatcgataaatatgtgaaggaaatgcctg
10723	>gene78.INS3	ctccttcttgcaattctcaacttcttcaacacttgagcggagtcgggtgc
10724	>gene78.m01INS4	caacactgcccagaggttcaatcgatgtgtgataggatcagtggtcaggg
10725	>gene78.mINS403	gaaggcgaggagactgctaataatcgataaatatgtgaaggaaatgcctg
10726	>gene78.INS4a	gaacgtatgacatgcgagagacgctgtagtggaaaaacccacgaagcg
10727	>gene78.INS4b	gaaaccgctgattatactgcggagaagggtgggtgagtataaagactatac
11345	>gene78.01a_40	aagtagatttgatttccgaaacgccttctcccgttctt
11346	>gene78.01b_40	accacaaatagtcctcaaaaatcacaagaaaactcaca
11347	>gene78.03a_40	gcagcgggtgtaaaaagtagtaaacgtggtaattaaaag
11348	>gene78.03b_40	gaaaactaatcaagggttaaacgtggatcccatggcaattc
11349	>gene78.m0103_40	cactgcccagaggttcaatcgataaatatgtgaaggaaat
11350	>gene78.m01INS3_40	ctgccagaggttcaatcgatccgatgatcctaataaggg
11351	>gene78.mINS303_40	gtatcgtccatcatagatcgataaatatgtgaaggaaat
11352	>gene78.INS3_40	tcttgcaattctcaacttcttcaacacttgagcggagtc
11353	>gene78.m01INS4_40	ctgccagaggttcaatcgatgtgtgataggatcagtggt
11354	>gene78.mINS403_40	cgaggagagactgctaataatcgataaatatgtgaaggaaat
11355	>gene78.INS4a_40	tatgagcatgcgagagacgctgtagtggaaaaacccacg
11356	>gene78.INS4b_40	cgctgattatactgcggagaagggtgggtgagtataaagac
11357	>gene78.01a_50_LNA3	mCctGaaAgtAgaTtGtAttTccGaaAcgmCctTctmCccGttmCttAagTc
11358	>gene78.01b_50_LNA3	mCatAtamCcamCaaAtaGtcmCctmCaaAaaTcamCaaGaaAacTcamCaamCacTg
11359	>gene78.03a_50_LNA3	GatTtgmcAgmCggTggTaaAaaGtaTgaAaamCgtGgtAatTaaAagGtcTc
11360	>gene78.03b_50_LNA3	mCcaAtgAaaActAatmCaaAggTaaAcgTggAtcmCcaTggmCaaTtcmCcgGg
11361	>gene78.m0103_50_LNA3	mCacAacActGccmCagAggTtcAatmCgaTaaAtaTgtGaaGgaAatGccTg



11362	>gene78.m01INS3_50_LNA3	mCaamCacTgcmCcaGagGttmCaaTcgAtcmCgaTgaTccTaaTgaAggmCgcmCc
11363	>gene78.mINS303_50_LNA3	GtcmCagTatmCgtmCcaTcaTagTatmCgaTaaAtaTgtGaaGgaAatGccTg
11364	>gene78.INS3_50_LNA3	mCtcmCttmCttGcaTtcTtcAacTtcmCttmCaamCacTtgAgcGgaGtcGgtGc
11365	>gene78.m01INS4_50_LNA3	mCaamCacTgcmCcaGagGttmCaaTcgAtgTgtGatAggAtcAgtGttmCagGg
11366	>gene78.mINS403_50_LNA3	GaaGgcGaaGgaGacTgcTaaTatmCgaTaaAtaTgtGaaGgaAatGccTg
11367	>gene78.INS4a_50_LNA3	GaamCgtAtgAgcAtgmCgaGagAcgmCtgTagTtgGaaAaamCccAcgAagmCg
11368	>gene78.INS4b_50_LNA3	GaaAccGctGatTatActGcgGagAagGtgGgtGagTatAaaGacTatAc
11369	>gene78.01a_40_LNA3	aAgtAgaTttGttAttTccGaaAcgmCctTctmCccGttmCtt
11370	>gene78.01b_40_LNA3	amCcamCaaAtaGtcmCctmCaaAaaTcamCaaGaaAacTcamCaa
11371	>gene78.03a_40_LNA3	gmCagmCggTggTaaAaaGtaTgaAaamCgtGgtAatTaaAag
11372	>gene78.03b_40_LNA3	gAaaActAatmCaaAggTaaAcgTggAtcmCcaTggmCaaTtc
11373	>gene78.m0103_40_LNA3	cActGccmCagAggTtcAatmCgaTaaAtaTgtGaaGgaAat
11374	>gene78.m01INS3_40_LNA3	cTgcmCcaGagGttmCaaTcgAtcmCgaTgaTccTaaTgaAgg
11375	>gene78.mINS303_40_LNA3	gTatmCgtmCcaTcaTagTatmCgaTaaAtaTgtGaaGgaAat
11376	>gene78.INS3_40_LNA3	tmCttGcaTtcTtcAacTtcmCttmCaamCacTtgAgcGgaGtc
11377	>gene78.m01INS4_40_LNA3	cTgcmCcaGagGttmCaaTcgAtgTgtGatAggAtcAgtGtt
11378	>gene78.mINS403_40_LNA3	cGaaGgaGacTgcTaaTatmCgaTaaAtaTgtGaaGgaAat
11379	>gene78.INS4a_40_LNA3	tAtgAgcAtgmCgaGagAcgmCtgTagTtgGaaAaamCccAcg
11380	>gene78.INS4b_40_LNA3	cGctGatTatActGcgGagAagGtgGgtGagTatAaaGac

### B. Printing and coupling of the splice isoform-specific microarrays

The splice variant capture probes were synthesized with a 5' anthraquinone (AQ)-modification, followed by a hexaethyleneglycol-2 (HEG2) linker. The capture probes were first diluted to a 5 20 µM final concentration in 100 mM Na-phosphate buffer pH 7.0, and spotted on the Immobilizer polymer microarray slides (Exiqon, Denmark) using the Biochip Arrayer One (Packard Biochip Technologies, USA) with a spot volume of 2x 300 pl and 300 µm between the spots. The capture probes were immobilized onto the microarray slide by UV irradiation in a Stratalinker with 2300 µjoules (Stratagene, USA). Non-immobilized capture probe 10 oligonucleotides were removed from the slides by washing the slides two times 15 min. in 1xSSC. After washing, the slides were dried by centrifugation at 1000x g for 2 min., and stored in a slide box until microarray hybridization.

### C. Construction of the Splice Variant Clones

The recombinant splice variant constructs were cloned into the Triamp18 vector (Ambion, 15 USA). The constructs were sequenced to confirm their construction. The plasmid clones were transformed into *E. coli* XL10-Gold (Stratagene, USA).

Genomic DNA was prepared from a wild type standard laboratory strain of *Saccharomyces cerevisiae* using the Nucleon MiY DNA extraction kit (Amersham Biosciences, USA) according to the supplier's instructions. Amplification of the partial yeast gene was done by 20 standard PCR using yeast genomic DNA as template. In the first step of amplification, a forward primer containing a restriction enzyme site and a reverse primer containing a universal linker sequence were used. In this step 20 bp was added to the 3'-end of the amplicon, next to

the stop codon. In the second step of amplification, the reverse primer was exchanged with a nested primer containing a poly-T<sub>20</sub> tail and a restriction enzyme site. The *SWI5* amplicon contains 730 bp of the *SWI5* ORF plus 20 bp universal linker sequence and a poly-A<sub>20</sub> tail.

The PCR primers used were;

5 YDR146C-For-EcoRI: acgtgaattcaaatacagacaatgaaggagatga

YDR146C-Rev-Uni: gatccccgggaattgccatgttacctttgattagttttcattggc

Uni-polyT-BamHI: acgtggatcctttttttttttttttttgatccccgggaattgccatg,

The PCR amplicon was cut with the restriction enzymes, *EcoRI* + *BamHI*. The DNA fragment was ligated into the pTRIamp18 vector (Ambion, USA) using the Quick Ligation Kit 10 (New England Biolabs, USA) according to the supplier's instructions and transformed into *E. coli* DH-5 $\alpha$  by standard methods.

C1. Construction of the recombinant splice variant #1 (Triamp18/swi5-rubisco)

The *Arabidopsis thaliana* Rubisco small subunit ssu2b gene fragment (gi17064721) was amplified from genomic DNA by primers named DJ305 5'-

15 ACTATGATGGACGATACTGGAC-3' and DJ306 5'-

ATTGGATCGATCCGATGATCCTAATGAAGGC-3', containing *ClaI* restriction site linkers.

The purified PCR fragment was digested with *ClaI* and then cloned into the *swi5* (gi:7839148) vector at the unique *ClaI* site (atcgat) giving each insert a flanking sequence from the original yeast *SWI5* insert (named exon01 and exon 03, see Figure 48). The product was inserted in the

20 reverse orientation, so that the insert sequence is:

atcgatCCGATGATCCTAATGAAGGCGCCCGGGTACTCCTTCTTGCATTCTTCAACTTC  
CTTCAACACTTGAGCGGAGTCGGTGCATCCGAACAATGGAAGCTTCCACATTGTCC  
AGTATCGTCCATCATAGTatcgat

25

Nucleotide sequence analysis revealed a difference between the sequence of *A. thaliana* Rubisco expected from the GenBank database and that obtained from all sequenced constructs and PCR products. Position 30 in the Rubisco insert is C rather than the expected A. This SNP was probably created by PCR. None of the oligonucleotide capture probes used in the example 30 cover this region.

Rubisco seq. in genbank TCCTAATGAAGGCGCCA

The sequence obtained

from the plasmid construct TCCTAATGAAGGCGCCC

C2. Construction of the recombinant splice variant # 2 (Triamp18/swi5-lea)

The *Arabidopsis thaliana* Lea gene (gi1526423) was amplified from genomic DNA with primers named DJ307 5'-GGAATTATCGATGTGTGATAGGATCAGTGTTTCAG-3', and DJ308 5'-AATTGGATCGATATTAGCAGTCTCCTTCGCC-3', including the *Cla*I linker sites as above. The PCR fragment was digested with *Cla*I cloned into the yeast SWIS IVT construct 5 as above at the unique *Cla*I site.

The fragment was inserted in the forward orientation, resulting in the following insert sequence:  
 atcgatGTGTGATAGGTTTCAGTGTTTCAGGGCTGTCCAAGGAACGTATGAGCATGCGAG  
 AGACGCTGTAGTTGGAAAAACCCACGAAGCGGCTGAGTCTACCAAAGAAGGAGCT  
 CAGATAGCTTCAGAGAAAGCGGTTGGAGCAAAGGACGCAACCGTCGAGAAAGCTA  
 10 AGGAAACCGCTGATTATACTGCGGAGAAGGTGGGTGAGTATAAAGACTATACGGT  
 TGATAAAGCTAAAGAGGCTAAGGACACAACCTGCAGAGAAGGCGAAGGAGACTGCT  
 AATatcgat.

Figure 48 shows the construction of the recombinant splice variants in the *in vitro* transcription 15 vector. The small bars show the location of the oligonucleotide capture probes used in this example. The sequences of the capture probes are shown in Table 15.

#### D. Preparation of target

##### D1. *In vitro* RNA preparation from splice variant vectors

*In vitro* RNA from the splice variants were made using the MEGAscript™ high yield 20 transcription kit according to the manufacturer's instructions (Ambion, USA). The yield of IVT RNA was quantified at a Nanodrop spectrophotometer (Nanodrop Technologies, USA)

##### D2. Isolation of total RNA from *C. elegans*

Strains and growth conditions: *C. elegans* wild-type strain (Bristol-N2) was maintained on nematode growth medium (NG) plates seeded with *Escherichia coli* strain OP50 at 20 °C, and 25 the mixed stages of the nematode were prepared as described in Hope, I. A. (ed.) "C. elegans – A Practical Approach", Oxford University Press 1999. The samples were immediately flash frozen in liquid N<sub>2</sub> and stored at – 80 °C until RNA isolation.

A 100 µl aliquot of packed *C. elegans* worms from a mixed stage population was homogenized using the FastPrep Bio101 from Kem-En-Tec for 1 min, speed 6 followed by 30 isolation of total RNA from the extracts using the FastPrep Bio101 kit (Kem-En-Tec) according to the manufacturer's instructions. The eluted total RNA was ethanol precipitated for 24 hours at – 20°C by addition of 2.5 volumes of 96% EtOH and 0.1 volume of 3M Na-acetate, pH 5.2 (Ambion, USA), followed by centrifugation of the total RNA sample for 30 min at 13200 rpm. The total RNA pellet was air-dried and redissolved in 10 µl of diethylpyrocarbonate (DEPC)- 35 treated water (Ambion, USA) and stored at – 80°C.

### E. Fluorochrome-labelling of the target

The following fluorochrome-labelled cDNA targets were synthesized to test the performance of 'merged' splice junction probes that encompass exon borders. Synthetic RNAs corresponding to three artificial splice variants; #1 (exon01-INS3-exon03 (1-INS3-3), #2 5 (exon01-INS4-exon03) (01-INS3-3) and #3 without the middle exon (01-03) were spiked into 10µg of *C. elegans* reference total RNA samples in various combinations and concentrations prior to fluorochrome-labelling with either Cy3 or Cy5 as indicated in Table 16. At the same time 10µg of *C. elegans* reference total RNA was labeled with Cy3 for control experiments. Hybridizations were performed with Cy3- and Cy5 labeled *C. elegans* RNA + spike RNA mix .

10 The details of RNA samples and synthetic RNA spikes are shown in Table 16. The RNA samples were combined in individual labeling reactions with 5 µg anchored oligo(dT<sub>20</sub>) primer and DEPC-treated water to a final volume of 8 µl. The mixture was heated at 70°C for 10 min, quenched on ice for 5 min, followed by addition of 20 units of Superscript RNase inhibitor (Ambion, USA), 1 µl dNTP solution (10mM each dATP, dGTP, dTTP and 0.4 mM dCTP, and

15 3 µl of Cy3-dCTP or Cy5-dCTP, Amersham Biosciences, USA), 4 µl 5 x RTase buffer (Invitrogen), 2µl 0.1 mM DTT (Invitrogen), 400 units of Superscript II reverse transcriptase (Invitrogen, USA) and DEPC-treated water to 20 µl final volume. Background hybridization to merged capture probes was monitored with 10µg of *C. elegans* reference RNA alone labeled with Cy3-dCTP; according to the labeling method described above for the splice variant spikes.

20 All cDNA syntheses were carried out at 42°C for 2 hours, and the reaction was stopped by incubation at 70°C for 5 min., followed by incubation on ice for 5 min.

Unincorporated dNTPs were removed by gel filtration using MicroSpin S-400 HR columns as described in the following: Pre-spin the column 1 min at 1500 x g in a 1.5 ml tube and place the column in a new 1.5 ml tube. Slowly apply the cDNA sample to the top centre of

25 the resin, spin 1500-x g for 2 min and collect the eluate. The RNA was hydrolyzed by adding 3 µl of 0.5 M NaOH, mixing and incubating at 70 °C for 15 min. The samples were neutralized by adding 3 µl of 0.5 M HCl and mixing, followed by addition of 450 µl 1xTE, pH 7.5 to the neutralized sample and transfer onto a Microcon-30 concentrator (prior to use, spin 500 µl 1xTE through the column to remove residual glycerol). The samples were centrifuged at 14000-

30 x g in a microcentrifuge for 12 min. Spinning was continued until volume was reduced to 5 µl. The labelled cDNA probes were eluted by inverting the Microcon-30 tube and spinning at 1000-x g for 3 min.

Table 16. Synthetic splice variant RNAs spiked into *C. elegans* samples\*.

Spike RNA concentration	Ratio	Splice variant RNAs	Observed ratio LNA 50mer	STDEV LNA 50mer	Observed ratio LNA 40mer	STDEV LNA 40mer	Expected ratio
1000 ppm	5	Cy3: spike 01-INS3-03	0.76	0.05	0.61	0.19	0.83
	1	Cy3: spike 01-03	0.24	0.05	0.39	0.19	0.17
	1	Cy5: spike 01-INS3-03	0.16	0.04	0.06	0.03	0.17
	5	Cy5: spike 01-03	0.84	0.04	0.94	0.03	0.83
1000 ppm	5	Cy3: spike 01-INS3-03	0.77	0.11	0.68	0.22	0.83
	1	Cy3: spike 01-INS4-03	0.23	0.11	0.32	0.22	0.17
	1	Cy5: spike 01-INS3-03	0.12	0.04	0.11	0.15	0.17
	5	Cy5: spike 01-INS4-03	0.88	0.04	0.89	0.15	0.83
1000 ppm	5	Cy3: spike 01-INS3-03	0.88	0.08	0.87	0.10	0.83
	1	Cy3: spike 01-INS4-03	0.12	0.08	0.13	0.10	0.17
	1	Cy5: spike 01-INS3-03	0.22	0.12	0.15	0.12	0.17
	5	Cy5: spike 01-INS4-03	0.78	0.12	0.85	0.12	0.83
100 ppm	5	Cy3: spike 01-INS3-03	0.89	0.15	0.11	0.08	0.83
	1	Cy3: spike 01-INS4-03	0.11	0.15	0.89	0.08	0.17
	1	Cy5: spike 01-INS3-03	0.48	0.2	0.57	0.31	0.17
	5	Cy5: spike 01-INS4-03	0.52	0.2	0.43	0.31	0.83
10 ppm	5	Cy3: spike 01-INS3-03	0.61	0.2	0.09	0.14	0.83
	1	Cy3: spike 01-INS4-03	0.39	0.2	0.91	0.14	0.17
	1	Cy5: spike 01-INS3-03	0.34	0.18	0.19	0.22	0.17
	5	Cy5: spike 01-INS4-03	0.66	0.18	0.81	0.22	0.83

\* Parts per million (ppm) calculations indicate spike transcripts per total transcripts in the hybridisation mix. Calculations are based on an average *C. elegans* RNA being 1000 5 nucleotides as in Hill *et al.* (2000). Science 290:809-812.

#### F. Microarray hybridization

The fluorochrome-labelled cDNA samples, respectively, were combined. The following was added: 3.75 µl 20x SSC (3x SSC final, pass through 0.22 µ filter prior to use to remove particulates) yeast tRNA (1 µg/µl final) 0.625 µl 1 M HEPES, pH 7.0 (25 mM final, pass 10 through 0.22 µ filter prior to use to remove particulates) 0.75 µl 10 % SDS (0.3 % final) and DEPC-water to 25 µl final volume. The labelled cDNA target samples were filtered in Millipore 0.22 µ filter spin column (Ultrafree-MC, Millipore, USA) according to the manufacturer's instructions, followed by incubation of the reaction mixture at 100 °C for 2-5 min. The cDNA probes were cooled at room temp for 2-5 min by spinning at max speed in a 15 microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was carefully placed on top of the microarray spotted on Immobilizer™ MicroArray Slide and the hybridization mixture was applied to the array from the side. An aliquot of 30 µL of 3xSSC was added to both ends of the hybridization chamber, and the Immobilizer™ MicroArray Slide was placed in the hybridization chamber (DieTech, USA). The chamber was sealed watertight and incubated at

65°C for 16-18 hours submerged in a water bath. After hybridisation, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The slides were washed sequentially by plunging gently in 2 x SSC/0.1% SDS at room temperature until the cover slip falls off into the washing solution, then in 1x SSC pH 7.0 (150 mM NaCl, 15 mM Sodium Citrate) at room temperature for 1 min, then in 0.2 x SSC, pH 7.0 (30 mM NaCl, 3 mM Sodium Citrate) at room temperature for 1 min, and finally in 0.05 x SSC (7.5 mM NaCl, 0.75 mM Sodium Citrate) for 5 sec, followed by drying of the slides by spinning at 1000 x g for 2 min. The slides were stored in a slide box in the dark until scanning.

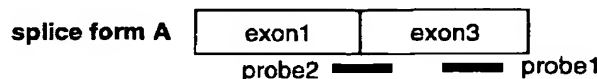
#### G. Microarray data analysis.

10 The splice variant microarray was scanned in a ScanArray 4000XL confocal laser scanner (Packard Instruments, USA). The hybridisation data were analysed using the GenePix Pro 4.01 microarray analysis software (Axon, USA). The *C. elegans* reference RNA alone converted to first strand cDNA and labelled with Cy3-dCTP did not produce significant fluorescence intensity signals from the LNA substituted spike RNA specific capture probes.

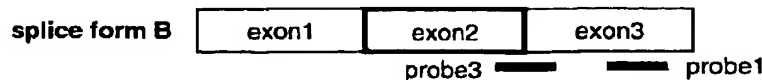
#### 15 G1. A mathematical formula for analysis of the microarray data for alternative splicing

One of the major limitations to comparative microarray hybridisation assays is that only identical probes can be compared between samples. Different alternative splice forms are detected using different probes, and this will tell directly if one splice form is more abundant in a given tissue compared to another. However, the estimation of the ratios between splice forms in a single tissue is not directly accessible. Given an example similar to that described below we employ the following calculations to calculate quantities of splice variants from array data. The theoretical justification is shown. To our knowledge this justification has not been used by any previous analysis.

25



30



The above scenario is tested in a comparative hybridisation, with two channels: I & II (signal from probe2 in channel I is called probe2(I), and so forth). Probe1 hybridises to both splice forms, Probe2 hybridises to A only, Probe3 hybridises to B only.

Since every transcript will hybridise to probe1, and every transcript will hybridise to either probe2 or probe3, there exists some relationship between the following:

probe1(I) ~ {probe2(I) and probe3(I)}.

probe1(II) ~ {probe2(II) and probe3(II)}.

5 For simplicity we assume that systematic differences between channels have already been eliminated through normalisation, although this is not essential.

We now imagine a factor (x) that will transform the signal of probe2 into a value directly comparable to probe1. Likewise we imagine factor (y) for probe3. As long as we are not facing saturation in the hybridisations, the assumption of a linear relationship between absolute probe

10 signals is reasonable.

The introduction of variables x & y will give the following equations:

probe1(I) = (x)probe2(I) + (y)probe3(I).

probe1(II) = (x)probe2(II) + (y)probe3(II).

Since all signals are measurable, the above is two linear equations with two unknown variables, 15 that can easily be solved. Further the ratio between (x)probe2(I) & (y)probe3(I) will provide the ratio between splice form A and B in channel I. Similarly, the ratio of (x)probe2(II) to (y)probe3(II) is used for channel II.

Data normalization is not required for this method.

In the above equations, probe1 denotes all probes that will hybridize to both 20 spliceforms, probe2 denotes probes that specifically will hybridize to spliceform A but not B, and probe3 denotes probes that will specifically hybridize to spliceform B but not A.

In the case where two spliceforms consist of gene78 with two different inserts/middle exons (INS3& INS4), probes can be grouped as follows (only LNA 40mers are considered here):

Probes that will hybridize to both constructs	Probes that will hybridize to INS3 constructs only	Probes that will hybridize to INS4 constructs only
Gene78.01a_40_LNA3	Gene78.INS3_40_LNA3	Gene78.INS4_40_LNA3
Gene78.01b_40_LNA3	Gene78.m01INS3_40_LNA3	Gene78.m01INS4_40_LNA3
Gene78.03a_40_LNA3	Gene78.mINS303_40_LNA3	Gene78.mINS403_40_LNA3
Gene78.03b_40_LNA3		

25

The equations can be solved with any combinations of one representative from each probe group. This gives a total of 48 (4x3x3) possible ways of solving the equations. The estimated quantities of the constructs are given as the average of all possible solutions (equations yielding 30 non-positive solutions are ignored). This was done for all comparative hybridizations. Note that

when comparing with gene78 with no insert, only 12 equations are possible (The, since the artificial splice variant construct with no insert has only one specific probe). The results from analysis of the microarray hybridization data from the RNA pools spiked with different splice isoforms at different ratios and concentrations are shown in Figures 49 and 50.

## 5 RESULTS

Figure 49 shows detection of alternatively spliced mRNAs using LNA-substituted 50-mer oligonucleotide capture probes in a bar diagram format. Figure 50 shows detection of alternatively spliced mRNAs using LNA-substituted 40-mer oligonucleotide capture probes. Both 50-mer and 40-mer LNA-DNA mixmer substituted oligonucleotide capture probes, 10 substituted with an LNA nucleotide at every third nucleotide position, were able to provide highly accurate measurements for fold-changes in the expression of three homologous, alternatively spliced mRNA variants in the concentration range of 1000 ppm to 10 ppm. The quantification of the splice isoforms was carried out using a set of both internal, exon-specific probes and merged, splice junction specific probes, printed onto microarrays and hybridized 15 with complex cDNA target pools spiked with different cloned artificial splice isoforms in which the middle exon was either alternatively skipped or excluded completely resulting in the three different splice isoforms; 01-INS3-03, 01-INS4-03 and 01-03.

### 20 Example 53. Expression Profiling of Toxicological responses in *Caenorhabditis elegans* using LNA Oligonucleotide Microarrays and beta-naphthoflavone and primaquine as model compounds.

The present patent example demonstrates the use of the Exiqon *C. elegans* LNA tox oligoarray in gene expression profiling experiments in the nematode *Caenorhabditis elegans*. The *C. elegans* tox oligoarray will monitor the expression of a selection of 110 genes relevant for 25 general stress response and for the metabolism of toxic compounds. Two different capture probes for each of these target genes have been designed, and included on the LNA tox array. In addition, the *C. elegans* LNA tox oligoarray contains capture probes providing control for cDNA synthesis efficiency and the developmental stage of the nematode. Capture probes for constitutively expressed genes for data set normalization is also included on the *C. elegans* 30 LNA tox oligoarray.

#### A. Cultivation of *C. elegans* worms

For all cultures the sample is divided into two, and one half of the sample is used as the control, the other as the treated sample. Worm samples are harvested and sucrose cleaned by standard methods. For heat shock treatment: the heat shock sample is added to S-media preheated to



33°C in a 1 L flask suspended in a water bath at 33°C, the other sample is added to a 1 L flask with S-media at 25°C. Both samples are shaken at approximately 100 rpm. for an hour. For  $\beta$ -naphthoflavone and primaquine treatment: 0.5 mL of 5 mg/mL  $\beta$ -naphthoflavone in DMSO or 0.5 mL of 20 mg/mL primaquine in DMSO were added to each 500 mL volume of S-media 5 culture after 28 hours of growth from L1. At the same time 0.5 mL of DMSO was added to the control. Incubation was for 24 hours. Samples are then harvested by centrifugation at 3000 $\times$ g suspended in RNALater™ (Ambion) and immediately frozen in liquid nitrogen.

#### B. RNA extraction

RNA was extracted from the worm samples using the FastRNA® Kit, GREEN (Q-BIO) 10 essentially according to the suppliers' instructions.

#### C. Design and synthesis of the LNA capture probes

Capture probes were designed using an in-house developed software. Regions with unique mRNA sequence of the selected target genes were identified. The optimal 50-mer oligonucleotide sequences with respect to T<sub>m</sub>, self-complementarity and secondary structure 15 were selected. LNA modifications were incorporated to increase affinity and specificity.

#### D. Printing of the LNA microarrays

The microarrays were printed on Immobilizer™ MicroArray Slides (Exiqon, Denmark) using the MicroGrid II from Biorobotics (UK). The arrays were printed with a 10  $\mu$ M capture probe solution. Two replicas of the capture probes were printed on each slide.

#### 20 E. Synthesis of fluorochrome labelled first strand cDNA from total RNA

15  $\mu$ g of *C. elegans* total RNA was combined with 5  $\mu$ g oligo dT primer (T20VN) in an RNase free, pre-siliconized 1.5 mL tube, and the final volume was adjusted with DEPC-water to 14.5  $\mu$ L. The reaction mixture was heated at +70°C for 10 min., quenched on ice 5 min., spin 20 seconds, followed by addition of 1  $\mu$ L SUPERase-In™ (20U/ $\mu$ L, Ambion, USA), 6  $\mu$ L 25 5 $\times$ RTase buffer (Invitrogen, USA), 3  $\mu$ L 0.1 M DTT (Invitrogen, USA), 1.5  $\mu$ L dNTP (20mM dATP, dGTP, dTTP; 4 mM dCTP in DEPC-water, Amersham Biosciences, USA), and 3  $\mu$ L Cy3™-dCTP or Cy5™-dCTP (Amersham Biosciences, USA). First strand cDNA synthesis was carried out by adding 1  $\mu$ L of Superscript™ II (Invitrogen, 200 U/mL), mixing and incubating the reaction mixture for 1 hour at 42°C. An additional 1  $\mu$ L of Superscript™ II was added and 30 the cDNA synthesis reaction mixture was incubated for an additional 1 hour at 42°C; the reaction was stopped by heating at 70°C for 5 min., and quenching on ice for 2 min. The RNA was hydrolyzed by adding 5  $\mu$ L of 1 M NaOH, and incubating at 70°C for 15 min. The samples were neutralized by adding 5  $\mu$ L of 1 M HCl, and purified by adding 450  $\mu$ L 1 $\times$ TE buffer, pH

7.5 to the neutralized sample and transferring the samples onto a Microcon-30 concentrator. The samples were centrifuged at 14000×g in a microcentrifuge for ~8 min, the flow-through was discarded and the washing step was repeated twice by refilling the filter with 450 µl 1×TE buffer and by spinning for ~12 min. centrifugation was continued until the volume was reduced 5 to about 5 µL, and finally the labelled cDNA probe was eluted by inverting the Microcon-30 tube and spinning at 1000×g for 3 min.

#### F. Synthesis of fluorochrome labelled cRNA from total RNA

First and second strand cDNA syntheses were made using the MessageAmp™ aRNA Kit (Ambion) according to suppliers' instructions. Five microgram of *C. elegans* total RNA was used as template for cDNA syntheses. Syntheses of fluorescent cRNA were made according to the MessageAmp™ aRNA Kit (Ambion) protocol with minor modifications. Cy3™-UTP or Cy5™-UTP (6 µl of a 5 mM solution Amersham Biosciences, USA) replaced biotin-CTP. The final concentration of ATP, CTP, and GTP was 7.5 mM whereas the concentration of UTP was reduced to 4.9 mM.

#### 15 G. Hybridization with fluorochrome-labelled cDNA or cRNA

The arrays were hybridized overnight using the following protocol. The Cy3™ and Cy5™-labelled cDNA or cRNA samples were combined in one tube followed by addition of 3 µL 20×SSC (3×SSC final), 0.5 µL 1 M HEPES, pH 7.0 (25 mM final), 25 µg yeast tRNA (1.25 µg/µL final), 0.6 µL 10% SDS (0.3% final), and DEPC-treated water to 20 µL final volume.

- 20 The labelled cDNA target sample was filtered in a Millipore 0.22 micron spin column according to the manufacturer's instructions (Millipore, USA), and the probe was denatured by incubating the reaction at 100°C for 2 min. The sample was cooled at 20-25°C for 5 min. by spinning at max speed in a microcentrifuge. A LifterSlip (Erie Scientific Company, USA) was carefully placed on top of the microarray spotted on Immobilizer™ MicroArray Slide and the
- 25 hybridization mixture was applied to the array from the side. An aliquot of 30 µL of 3×SSC was added to both ends of the hybridization chamber, and the Immobilizer™ MicroArray Slide was placed in the hybridization chamber. The chamber was sealed watertight and incubated at 65°C for 16-18 hours submerged in a water bath. After hybridisation, the slide was removed carefully from the hybridization chamber and washed using the following protocol. The Lifterslip
- 30 coverslip was washed off in 2×SSC, pH 7.0 containing 0.1% SDS at room temperature for 1 min., followed by washing of the microarrays subsequently in 1.0×SSC, pH 7.0 at room temperature for 1 min, and then in 0.2×SSC, pH 7.0 at room temperature for 1 min. Finally the slides were washed for 5 seconds in 0.05×SSC, pH 7.0. The slides were then dried by

centrifugation in a swinging bucket rotor at approximately 200 G for 2 min. The slide is now ready for scanning.

#### H. Data analysis.

Following washing and drying, the slides were scanned using a ScanArray 4000XL scanner 5 (Perkin-Elmer Life Sciences, USA), and the array data were processed using the GenePix™ Pro 4.0 software package (Axon, USA). The data in each image was normalized so that the ratio of means of all of the features is equal to 1.

## RESULTS

- 10 Use of LNA-modified oligonucleotide capture probes in Exiqons *C. elegans* LNA tox oligoarray clearly allows the identification of distinct expression profiles for *C. elegans* genes relevant for general stress response and for the metabolism of toxic compounds.

Table 17. Expression profiling using LNA Oligonucleotide Microarrays.

- 15 Log2 transformed fold of changes for selected genes in the two expression profiling experiments hybridised with cRNA target.

Gene name	Tox compound	
	Primaquine	beta-naphthoflavone
ABC_C34G6.4		1.01
ABC_F57C12.4	-1.11	
CYP_C03G6.15		2.35
CYP_C06B3.3		2.47
CYP_C49G7.8		2.40
CYP_F14F7.2	-1.24	-1.03
CYP_K07C6.5		2.68
CYP_K09D9.2		2.14
DC_W05G11.3	1.16	
ER_26S	-1.09	-1.01
HSP_C47E8.5		1.17
HSP_F26D10.3		1.05
HSP_F43D9.4		1.27
NAP_D2096.8		1.14
PPGB_F13D12.6	1.08	1.21
RAD_Y116A8C.133		1.13
RPL_K11H12.2	1.15	1.42
Ubi_F25B5.4		1.37

Table 18. LNA-modified oligonucleotide capture probes. LNA modifications are depicted by uppercase letters in the sequence, mC denotes LNA methyl cytosine.

Oligo Name	ence
CEABC_C34G6.4_u293_LNA3	TgcmCatTgcAcgGgcActTgtTcgAtcTccTtcTgtTttActTttGgaTg
CEABC_C34G6.4_u375_LNA3	TcaTtcTagGatTgcmCagAtgGttAtgAtamCtcAtgTcgGagAgaAagGa
CEABC_F57C12.4_u15_LNA3	mCcaAtgTtgTttAatTggTtgTaaTgtmCttGatGacmCtgmCatAatmCatAt
CEABC_F57C12.4_u480_LNA3	mCacAagAtcmCtgTgtTgtTctmCcgGaamCaaTgaAaaTgaActTagAtcmCa
CEABC_F57C12.5_u111_LNA3	TacTtgTtcTcgAcaAagGttGtgTagmCcgAgtTtgAcamCtcmCgaAgaAa
CEABC_F57C12.5_u444_LNA3	TgaActTggAtcmCctTctTtgmCatTtaGcgAtgAtcAaaTttGggAagmCg
CEABC_K08E7.9_d8_LNA3	TcaTtaAttTgtTgtAgcTttmCttTctmCgaTttTtgmCacGatmCttTccmCc
CEABC_K08E7.9_u51_LNA3	AggGtgmCctActAcaAacTgamCccAaaAgcAgaTgamCcgAgaAgaAatAa
CEABC_Y39D8C.1_u37_LNA3	AttGaaAgcGacGcgGaaAgtGccAlgTatTtcTaaTttTgtTttmCttTa
CEABC_Y39D8C.1_u422_LNA3	TtgTcaGcaTatmCaaGagTagAtaTggAagTggAtamCacTctGctAatmCc
CEADH_H24K24.3a_d3_LNA3	mCacmCttAttGcgTtcAatTttTgtTtcmCacmCtcmCtcmCtcmCgaAtamCgtTg
CEADH_H24K24.3a_u50_LNA3	TcamCaaGggAgaGagTctGcgGtcGgtGctGgcGttmCgaGaaAatAtaAc
CEAPEX_R09B3.1_u191_LNA3	mCatGcaTccmCgamCgaGaaGaaGtamCtcAttTtgGagTtaTctGgcGaaTt
CEAPEX_R09B3.1_u37_LNA3	GacmCatGctmCcgGtcGtcAtgmCaaAtcGacTtcTaaAttGctTctGatTa
CEAPO_C35D10.9_u15_LNA3	TtgmCatGctGttAaaAccTatmCgtGtamCaaTatTgcmCtgTatAttmCccmCt
CEAPO_C35D10.9_u609_LNA3	TggmCacAgcTtaAtaAcaAatTggAaaGtcGagGatTagTcgGtgTtgAa
CEAPO_C48D1.2_u176_LNA3	GacAcamCgcAaaGgaTatGgaTgtTgtTgaGctGctGacTgaAgtmCaaTa
CEAPO_C48D1.2_u23_LNA3	AgcAcgAaamCtcTgcmCgtmCtaAaaTtcActmCgtGatTcaTtgmCccAatTg
CEAPO_F20C5.1_u453_LNA3	AtgGtcAtamCtcTaaAatGggmCagAacTtcAacmCaaAtcAttmCtcGtcAg
CEAPO_F20C5.1_u96_LNA3	AacmCcgAgcTtgmCcgmCaaAgtGcaAgaAaaTtaTagAacGaaTgaAacAg
CEATPase_B0365.3_u31_LNA3	GgaTggGtcGagmCgtGagAccTacTaaAgaAcaGctTgtGaaTctTt
CEATPase_B0365.3_u386_LNA3	mCaamCgtTctmCgaTtcmCtcmCggAcaAgaAtgGacmCtaTgcmCaamCagA
CEATPase_C17H12.14_u356_LNA3	aaGa
3	TgcTcgTtaTccAgcTatTttGaaGggActTgtmCatGcaAggActTctTc
CEATPase_C17H12.14_u89_LNA3	mCcgTttAgaGctTatTgcTaamCcaGatTgtmCccAcaAgtmCagAacAgcTc
CEATPase_F55F3.3_u215_LNA3	TgamCggAcgmCtcmCtcmCccAlaTgtAttTgtTccAtcTtcmCcaGcaAccAa
CEATPase_F55F3.3_u275_LNA3	AgcTacTtcAttmCgamCaaGgaAcaTctmCggAaaAgtmCaaGtamCalmCccG
CEATPase_Y49A3A.2_u103_LNA3	AaaTtcAagGatmCcaGttGccGatGgtGaaGccAagAttmCgcAagGatTa
CEATPase_Y49A3A.2_u272_LNA3	mCgaTcgTttmCtgmCccAttmCtcmCaaGacTgtmCggTatGctmCaaGaaTatG
CECALR_Y38A10A.5_u238_LNA3	a
CECALR_Y38A10A.5_u296_LNA3	TcaGgaAcgAtcTttGacAacAttAtcAtcAccGacTctGttGagGagGc
CECAT_Y54G11A.5b_u137_LNA3	TgaActmCtcmCtcTtaTgaAagmCtgGggAgcmCatmCggAttmCgaTttGtgGc
CECAT_Y54G11A.5b_u189_LNA3	GaamCttTgcAggGccGctmCggGgaAtgTcaTgaTttmCatTatTaaGggAa
CECC_C03D6.3_u275_LNA3	GtcAatTctGggAgaAggTgtTggAtamCcgGggmCtcGggAgaGaaTgtGc
CECC_C03D6.3_u430_LNA3	AtgTaaAgaAggAatGctTccmCgaAtgGatTggAtaTttAttTgtmCcaGa
CECC_C07G2.3_d9_LNA3	GgamCcgAaaTttGtgmCagmCatGtcGgamCacGaaAttGatGgtmCtcAttTt
CECC_C07G2.3_u44_LNA3	mCagAcamCgaAggTtamCgaTagAtaAccAtcTctmCaaAgtmCtaTcgAccTc
	mCgamCgaTgtGcgTgtTccTgamCgaTgaAagAatGggAtaTtaAgaAaamCc

CEEC\_Y46G5A.2\_u331\_LNA3  
 CEEC\_Y46G5A.2\_u385\_LNA3  
 CEECoA\_C29F3.1\_u316\_LNA3  
 CEECoA\_C29F3.1\_u392\_LNA3  
 CEECoA\_F08A8.4\_u1094\_LNA3  
 CEECoA\_F08A8.4\_u1260\_LNA3  
  
 CEECoA\_F59F4.1\_u109\_LNA3  
 CEECoA\_F59F4.1\_u424\_LNA3  
 CEECoA\_Y25C1A.13\_u115\_LNA3  
 CEECoA\_Y25C1A.13\_u451\_LNA3  
 CEECOL\_C27H5.5\_u493\_LNA3  
 CEECOL\_C27H5.5\_u680\_LNA3  
 CEECOQ\_ZC395.2\_u199\_LNA3  
 CEECOQ\_ZC395.2\_u400\_LNA3  
 CEECRYZ\_F39B2.3\_u171\_LNA3  
 CEECRYZ\_F39B2.3\_u222\_LNA3  
  
 CEECyclin\_R02F2.1a\_u24\_LNA3  
 CEECyclin\_R02F2.1a\_u312\_LNA3  
 CEECyclin\_ZC168.4\_u203\_LNA3  
 CEECyclin\_ZC168.4\_u273\_LNA3  
 CEECYP\_B0213.15\_u133\_LNA3  
 CEECYP\_B0213.15\_u202\_LNA3  
 CEECYP\_B0304.3\_u38\_LNA3  
 CEECYP\_B0304.3\_u89\_LNA3  
 CEECYP\_C03G6.14\_u706\_LNA3  
 CEECYP\_C03G6.14\_u768\_LNA3  
 CEECYP\_C03G6.15\_d9\_LNA3  
 CEECYP\_C03G6.15\_u148\_LNA3  
 CEECYP\_C06B3.3\_u102\_LNA3  
 CEECYP\_C06B3.3\_u474\_LNA3  
 CEECYP\_C12D5.7\_u399\_LNA3  
 CEECYP\_C12D5.7\_u65\_LNA3  
 CEECYP\_C45H4.17\_u27\_LNA3  
 CEECYP\_C45H4.17\_u598\_LNA3  
 CEECYP\_C45H4.2\_u110\_LNA3  
 CEECYP\_C45H4.2\_u429\_LNA3  
 CEECYP\_C49C8.4\_u363\_LNA3  
 CEECYP\_C49C8.4\_u883\_LNA3  
 CEECYP\_C49G7.8\_d6\_LNA3  
 CEECYP\_C49G7.8\_u795\_LNA3  
 CEECYP\_F01D5.9\_u374\_LNA3  
 CEECYP\_F01D5.9\_u46\_LNA3

TtgTgcTccAtcGctGctmCcgmmCttAcaGacTtgAcaAcgmmCtcAccTttGc  
 AatGagmCggTtgTgcmCgtGtgAcgTcamCttmCgtmCacAgtGttGctmCtmmCt  
 AaaTtgAcamCcaAtcAaaTctGtcTcaTctmCctGagGacmCgtmCaamCttmCg  
 AatmCttTgtGtamCggAgaTggGgcAaaAggmCagmCaaGaaAgtAaamCcaAg  
 AggAcaAggGgcActActGgcAcaGgcTttGatTatTgcAgtGagAtaTt  
 TtaAtgGagGtgAcaAtgGgtTccTtgGatTcgAtaAatTccGagTgcmCc  
 GctmCttmCtcmCagTggGctmCaaAatAgtmCaamCtcAacAgaTcgGaaGttmC  
 t  
 AaaGctTcgAgaTggmCacGttmCgtmCtgTatmCtcGtgAagAacTtaTgmCa  
 GatTgcmCtgAacTttAtcAagAcgTggAatAtgAgcmCagmCtcmCtgTcgAc  
 GatmCttAtcAccGcgTgcGatAttmCgaGtaGctTcamCagGatGcgAttTt  
 GgaAagGaaGgaTccAttmCtcAgcTctGcamCttmCcamCcaTcaGagmCcaTg  
 TggAtamCaaGgaGggAtcTggmCagTggTggAtcTggAagTggTggAtaTg  
 TtgAaaGaamCtcmCttGccGacGatmCctGaaAcamCacAaaGaaTtgmCtgAa  
 AtgTggGatGagGagAaaGaamCatTtaGatAcaAtgGaaAgaTtaGctGc  
 AggmCtgAgcTctTggActTtgGcaTcaAcaTtgTctmCatTctTgaAggAa  
 TtaTggTtamCagAagGagmCtgTttAcgGtgTagmCatTggGaaTgtmCttmCc  
 mCacTtcAacmCaamCtcmCgtGttAatmCaaGcaAgcmCgcmCacmCatmCta  
 AtgAg  
 TctmCatTgcTcgTcgAggmCtmmCcaAcaAacActGgcAatAccmCaaTtaAt  
 TaaGaaAgtmCatTgaGgaTgcTgtmCgcTttGctmCgcmCgaAgtmCtcGtaTa  
 AagTtcAtcmCtgTtgAcgGaaTcgAggmCggAgaAtgmCtgTatmCggTcaTt  
 AcaGgaAatAtgAttTtgGatTtcGatTttGaaTcgGttGgtGctGccmCc  
 GctGagmCtgTatTtgGctAggAaaAtgTgtGttTttGatActTtaAatGa  
 AcgAggTttGgaTcamCaaTcaGaaTtcTgtGaaAtaAgcGttTttTggGa  
 AgtTctmCggTctAacAgtGtcTccmCgtTgaAtaTtcTtgTaaAatmCacAc  
 AtgAccActmCaaAatActGctAaaAgaTttGcaGcgGcaGaaGccGttAa  
 TtgAtaTggmCtgTatmCtgTatGgtTttTgaGgamCgtTttTgaGgaGtcGa  
 AttTatTcaTtcAtcmCatGtaAacTgtAtaTttTgaAttTgtGttGtaAa  
 GccAaaGcaGaaTtgTatTtgAtcTtcGgtAacmCttmCtcmCttmCgcTacAa  
 AttTtgAatmCttmCtgGgaAaaTgcmCatmCcamCtcGagAaamCgtTtcmCgtTt  
 mCtaAcgGagGatmCtcGccAatTatmCttTgaGagAcaAaamCtgAaamCtcmCt  
 AtcTagTccmCaaTgaAtcTccmCacAtgmCtgTtamCtcGtgAtgTtcAacTc  
 TttTgcTttmCatmCgcAaaAgcTcaAgaTtamCacAtgTcaGgtmCaaGccAa  
 mCcgmmCgamCttTaaAgaGaaGatmCatAaaTttGcaTtgTttTgtTgtAt  
 mCgaGggTgaTtcGgaGacTttmCagTaaTgtmCcaActTtcAaaTgtTgtmCa  
 TagAtamCaaGatAcaTccmCtcAaaAgaAggmCctAccGtcAatGgcmCaaAg  
 TcaAcgmmCgtmCtaTaaAtgAatmCacAacGagGtaTcaAcaTtcTccmCccTg  
 AtgmCtgAtgTtgAaaTtgmCtgGctAccGtaTtcmCaaAagAtamCtgTaaTc  
 AtgAatmCcaTggmCttGgamCatmCtcmCcgTttTtcAagGgaTatAaaAatGt  
 AtgmCaamCgaAttAggAaaAaaTtcAtcmCtgGaaTaaAaaAtaAttmCtaAa  
 AtcGctAcgAcaAtcTttmCcgAtgmCctTcgAagTttmCgaAagmCttCtcmCt  
 GagGtcGgtGgaGgaGgaAgtGgaAatTgamCggmCaaAatmCctGccmCaaGg  
 mCccTctTtgGgaTttmCcamCtcAagTttActGttmCggmCagmCagTgaTatAa

CECYP\_F08F3.7\_u25\_LNA3  
 CECYP\_F08F3.7\_u401\_LNA3  
 CECYP\_F14F7.2\_u397\_LNA3  
 CECYP\_F14F7.2\_u68\_LNA3  
 CECYP\_F42A9.5\_u435\_LNA3  
 CECYP\_F42A9.5\_u55\_LNA3  
 CECYP\_K07C6.3\_u3\_LNA3  
 CECYP\_K07C6.3\_u354\_LNA3  
 CECYP\_K07C6.4\_u118\_LNA3  
  
 CECYP\_K07C6.4\_u87\_LNA3  
 CECYP\_K07C6.5\_u7\_LNA3  
 CECYP\_K07C6.5\_u99\_LNA3  
 CECYP\_K09A11.3\_u362\_LNA3  
  
 CECYP\_K09A11.3\_u48\_LNA3  
 CECYP\_K09A11.4\_u238\_LNA3  
 CECYP\_K09A11.4\_u68\_LNA3  
 CECYP\_K09D9.2\_u151\_LNA3  
 CECYP\_K09D9.2\_u866\_LNA3  
 CECYP\_T10B9.10\_u410\_LNA  
 CECYP\_T10B9.10\_u56\_LNA  
 CECYP\_T10B9.7\_u102\_LNA3  
 CECYP\_T10B9.7\_u267\_LNA3  
 CECYP\_T19B10.1\_u100\_LNA3  
 CECYP\_T19B10.1\_u319\_LNA3  
 CECYP\_Y49C4A.9\_u121\_LNA3  
 CECYP\_Y49C4A.9\_u413\_LNA3  
 CECYP\_ZK177.5\_u394\_LNA3  
 CECYP\_ZK177.5\_u445\_LNA3  
 CEDAO\_C47A10.5\_d9\_LNA3  
 CEDAO\_C47A10.5\_u269\_LNA3  
 CEDC\_C01A2.3\_u373\_LNA3  
 CEDC\_C01A2.3\_u96\_LNA4  
 CEDC\_C34F6.1\_u301\_LNA3  
 CEDC\_C34F6.1\_u450\_LNA3  
 CEDC\_F33D11.3\_u126\_LNA3  
 CEDC\_F33D11.3\_u14\_LNA3  
  
 CEDC\_F46E10.2\_u392\_LNA3  
 CEDC\_F46E10.2\_u54\_LNA3  
 CEDC\_F56G4.2\_u382\_LNA3  
 CEDC\_F56G4.2\_u82\_LNA3  
 CEDC\_M162.2\_u103\_LNA3

GagTtgGttmCcamCagAatGctTagGacGttTaaAttmCgtmCacAaamCttTt  
 mCaaTatGgtTccmCatTttAgcAacTcaTatGaamCacAgaAgaTgtmCctTg  
 GaaAaaGgcGtcGacAttTtaTgtGacAcgTggAcamCttmCacTatGacAa  
 TaaTtgAatTacGggTctTttGlamCatAttAatTttAgtAtamCttTgtGa  
 AtaTcaAtgmCcamCtaTtaAtgAatmCacAacGtcTgtmCcaAtcTtcTccmCg  
 GgaGtgActAtgAaaGcaAagAgTatcmCgaTtgAaamCtgAaaGacAgamCa  
 AatmCttTaaTgaTaaTttAtgGgaTctGtaTttmCtcTttmCtgTcaAtaAa  
 AtgAgcmCcamCaaAtgTaaAagGatAcgAgaTtgAttmCggGaamCagTcaTg  
 AtcmCtgmCgaTatGacAttAagmCcamCatGgtTctGaamCctTcaAcaGaaGa  
 mCtgAacmCttmCcamCagAagAtaAacTtcmCgtAtaGcgmCtgGaaAaamCtc  
 mCt  
 AttTaaAggAatTcamCagmCtcAaaAaaTaaTaamCtamCcgGttmCagAgaTt  
 AatTtgAgcmCacAtgGcaAgTatmCaamCagAggAgamCaaTgcmCgtAcaGt  
 TgamCatTctActTaaAggGaaGaaAatAccAacTggTatcmCctTgtAttTg  
 TcamCcamCaaAgcmCatAcaTatGcgAgcTagTtcmCtcAggmCtgmCttAaam  
 Cc  
 TtcGacAaaActAttTtgGaaAgaAcaAtcmCcaTtcAgtGtcGgcAaamCg  
 TctGacAacAaaGccAtamCacGtgmCcgActAatTccAcaAtcAgcTagAa  
 TtgGcaAaaGcaGaaTtgTatTtaAtcTttGgaAacmCtcmCttmCttmCgcTa  
 TgaAtcTttmCaaActTatmCacTccTttTaaTacTatcmCgtTccTgtTtgGa  
 AttGagAttGtaTccAttGgcGtcTctTgtTcamCaaTcgAaaAtgTctmCa  
 AacTgcTacTatTgcGccAtcAagTgtGctGctmCaaActTaaAtcmCagGt  
 TtgAgamCagGaaAtaAgamCtaGaaTtcmCttTgaAacTggTggGaaGtgmCt  
 AagAtgTcaAagAatTcaAgcmCagAacGatGgtmCcamCcgAcgAgcmCatTa  
 AttGaamCcaActmCtgAaaTatAatGacAcaAaamCcaTgtmCtgGaaGtgGt  
 GgcAatGtgAcaAtaTctmCcaAtgGttmCttmCacAgcAatmCcatmCacGtgTt  
 mCtaTtcAatmCgaTatTttAtcAcamCcaTccAgtGctGgamCctmCcaTcaTt  
 GtcTcaGagAtgTgtAaaTttActTccmCtgmCaaTttGttTcamCgcAacTa  
 TtcmCgaAtgTttmCcaAttGggActGaaGttTcaAgaGtcAccmCagAaaAa  
 GatmCcaGcaTctTccAagmCttAcaTtcmCtcmCgtGctTgtAtcAagGaaAc  
 TttGaaAacmCtgTttTatTatTaaAatAgaTaaTtgAttAgTctGtamCg  
 AtamCgtTgcActGcaTccGgcTatGagGgaGccAaaAatmCttAggGgaGt  
 GcamCttmCcaTtcAtcTctGcaGctActAtgGctTtgGtgAcaAaaGttGg  
 mCcgTccAaaAgaAtgmCcaTctmCacAagTctTgaAatmCttAtaAagGtaGt  
 GagGgaTcaAcaGtaAccTcgTgcGgtAttGacAagGgaTgtmCcgGaaGg  
 GatGgtTctTcgAtcGcaAacAaaAcaGatGtgmCtcmCatTtamCatAcgGa  
 AtgGagAaaAtgGatmCtgAtgGagTtgmCagGaaGtgAtgGagmCtcmCagGa  
 TgaAtcTccAtaAatTatTcaAtgTttmCcaAatAttTaaTttAtcAatTg  
 GctmCaamCacGgtAggAtcmCtaTggAacmCgtmCggAggAgcAggmCctmCg  
 gAg  
 mCgtGacAacmCtcTtaTttAttTctGtaAaamCtgAttmCgcmCaaActTttGt  
 GaaGctTtcAaamCcaAatGagTtcmCttmCccGgaAtcmCcaAagAatAccAa  
 AcaAtgAaaAgaGagGatGgaAagGaaAtcGaaGtcTctGttmCttGacGa  
 GatGagGtamCatAacTttGtgTgcAgTatAggmCcaTctAcaGtamCctGc

CEDC\_M162.2\_u480\_LNA3 TtcmCatmCatmCacTaamCcgAttGtcmCtgAcaTtgAtgGccAaamCcaGggAa  
 CEDC\_R10E4.11\_u274\_LNA3 TcamCatTatmCgaAcaAgtActAgtAagmCatGctGtgAtgGagTgcmCgcTa  
 CEDC\_R10E4.11\_u397\_LNA3 mCacGgaGatmCacGacAtcAaaGcgGatTgcTtaGagTgtGgaAacmCgtmCt  
 CEDC\_T04C9.1\_u321\_LNA3 ActAtcTacGtgGcamCgtTggActmCatmCatmCgaTggGaamCgamCgtAtaAg  
 CEDC\_T04C9.1\_u64\_LNA3 TctmCtgGccAgtTcamCttTgtGatmCaaTctmCagAttmCgtmCcamCacAagAt  
 CEDC\_W02A2.3\_u32\_LNA3 mCtamCttmCcgCaaGaaGgcmCcgTcgTttmCtaAtcGatmCgaAcaTctmCac  
 Ac  
 CEDC\_W02A2.3\_u374\_LNA3 AtgGatGatmCgamCccActTgcmCacTgamCccAcaAtcmCcgCacTcamCta  
 CEDC\_W05G11.3\_u153\_LNA3 mCc  
 CEDC\_W05G11.3\_u51\_LNA3 AagAcgGagAggmCtgGagAgaAcgGtamCcgAtgGagAgcmCagGaamCtgAt  
 CEDC\_ZK863.5\_u256\_LNA3 mCcamCccAggAggAggGatAcaAgaGaaGaaAgtAcaGatTctmCcaActAa  
 CEDC\_ZK863.5\_u324\_LNA3 AgtTtcAcamCttmCttTttGccGttTtgGttmCccGttAtcAatmCcaTtgAt  
 CEEPFX\_Y55B1BR.4\_u161\_LNA3 mCttTtaTatTctmCatmCaaTttGttTccTacTtgGtcAgcTgaGgaTcgTt  
 CEEPFX\_Y55B1BR.4\_u93\_LNA3 TtcGgcAcaAatGgaGcaAaaGtaTcgTggTtaTtgTgaTgcGatTatTc  
 CEER\_18S\_u388\_LNA3 mCtamCtaTgaAtgAgcTcamCtgGacTcaTttAtcAacTcgAgtmCaaAagmCc  
 CEER\_18S\_u82\_LNA3 GttGgcGaaTctTcgGgtTcgTatAacTtcTtaGagGgaTaaGcgGtgTt  
 CEER\_26S\_u342\_LNA3 GaamCtgAttmCgaGaaGagTggGgamCtgTcgmCttmCgaGgtTtaAcgActTc  
 CEER\_26S\_u38\_LNA3 TgtTatTgcGaaAgtAatmCctGctTagTacGagAggAacAgcGggTtcAa  
 CEFOXO\_R13H8.1b\_u331\_LNA3 TgcAtamCgamCttGgtmCtcTtgGtcAagGtgTtgTatTcaGtaGagmCagTc  
 CEFOXO\_R13H8.1b\_u393\_LNA3 TgtGctmCagAatmCcamCttmCttmCgaAatmCcaAttGtgmCcaAgcActAacTt  
 Gt  
 CEGAPDH\_K10B3.7\_u21\_LNA3 TtaAgamCggAacmCaaTtgmCtcmCacmCacmCatmCatAccAcgAgtTgaAca  
 CEGAPDH\_K10B3.7\_u727\_LNA3 Gt  
 CEGBA\_F11E6.1a\_u232\_LNA3 AcaTtgmCtamCcaAggmCctAagmCcgCgtmCaaAttmCtcTaaGtcTgaAatG  
 CEGBA\_F11E6.1a\_u451\_LNA3 a  
 CEGLU\_C02A12.1\_u264\_LNA3 GttGagTccAccGgaGtcTtcAccAccAtcGagAagGccAatGctmCacTt  
 CEGLU\_C02A12.1\_u55\_LNA3 AgtAaaTtcmCttmCcamCgtGgaTctActmCgtGtgTtcAcaAagAtcGagGg  
 CEGLU\_C46F11.2\_u271\_LNA3 GgtmCcaAtaAtgGgaGacTggTtcmCgcGcaGaaAgtTatGcaGatGatAt  
 CEGLU\_C46F11.2\_u45\_LNA3 AgaAaamCttmCgtTggAccmCtgmCtaAggAgaAgtAttTcaAgcTtcTgaGc  
 CEGLU\_F26E4.12\_u109\_LNA3 GagmCacmCcgAagmCtcAagmCcaTatTtgGaaAcaAgamCcaTacTctTcaAa  
 CEGLU\_F26E4.12\_u480\_LNA3 GttAccmCtcTacAaaTctmCgcTtcAatmCcaAtgTtgTtcGcaGtcAccAa  
 CEGLU\_F26E4.12\_u109\_LNA3 mCcgAagAgcTcgTtamCtaTgcGagGagGtgTgaAgcmCggAatAatTttTt  
 CEGLU\_F26E4.12\_u480\_LNA3 AagTtcTtgGttGgamCgcGatGggAaaAttAtcAagAgaTttGgamCcaAc  
 CEGLU\_R07B1.4\_u166\_LNA3 AcgAttTcaAcgTcaAaaAtgmCtaAtgGtgAtgAcgTgtmCacTttmCggAt  
 CEGLU\_R07B1.4\_u38\_LNA3 AccTggGttGatGttTttGcgGctGaaAgtTtcTccAagmCtcAttGatTa  
 CEGLU\_T09A12.2\_u220\_LNA3 GaaGtamCgtmCtcmCcaAagAaaAgcTacmCccAgcTtaAggmCatTgcAcaAt  
 CEGLU\_T09A12.2\_u335\_LNA3 GcgCcaGatAtgTatTcaAagAtcGagGtaAatGgtmCagAacActmCatmCc  
 CEGLU\_T28A11.11\_u299\_LNA3 AatmCtamCagGgaAaaAggAttTcgAgtTgcmCgcGttTccAtgmCaaTcaAt  
 CEGLU\_T28A11.11\_u54\_LNA3 AgaTggmCaaAgaAgcAtamCatAacTgaAacTctTccmCggGgaGctActAc  
 CEGPD\_B0035.5\_u256\_LNA3 TgaAtaAacGggmCcgAacTaaAtcmCatTcgTcaGtgGaaAtgGgaAacAa  
 CEGPD\_B0035.5\_u478\_LNA3 GtcmCgtmCttmCctGatGctTatGaamCgcmCtaTttmCtcGaaGtaTtcAtgGg  
 CEHSP\_C09B8.6\_d8\_LNA3 TgtGgaAaaGctmCtcAacGagAagAaaGcaGaaGttmCgtAtamCaaTtcAa  
 CEHSP\_C09B8.6\_u286\_LNA3 AlaTcgmCcgCctGctTccTcamCcaAccmCgaAtaAcgmCaamCaaAaamCtt  
 Ta  
 AagAgcmCcamCtcAtcAagGatGaaAgtGatGgaAagActmCttmCgtmCtcAg

CEHSP\_C12C8.1\_u127\_LNA3 mCaaGatAttTtaAcaAaaAtgmCatmCaamCaaGaaGccmCaaTcaGgtTccGg  
 CEHSP\_C12C8.1\_u1531\_LNA3 mCttGggmCafTctGtamCggGatGctGtcAttActGtgmCctGcaTatTttAa  
 CEHSP\_C47E8.5\_u310\_LNA3 AagAagmCatmCtcGaaAtcAacmCcaGacmCacGctAtcAtgAagAcamCttmCg  
 CEHSP\_C47E8.5\_u361\_LNA3 AtgAaaGctmCaaGctmCttmCgtGatTccTctActAtgGgaTcaAtgGccGc  
 CEHSP\_F26D10.3\_u276\_LNA3 TtaAgcAgamCcaTtgAggAcgAgaAgcTcaAggAtaAgaTcaGccmCagAa  
 CEHSP\_F26D10.3\_u397\_LNA3 mCgtmCttTccAagGatGacAttGaamCgcAtgGtcAacGaaGctGagAaaTa  
 CEHSP\_F43D9.4\_u169\_LNA3 GtcGacTtgGctmCacAtcmCacAccGtcAtcAacAagGaaGgamCagAtgAc  
 CEHSP\_F43D9.4\_u275\_LNA3 mCaaTctTgaGggAcamCgtTctmCacmCatTgaGggAcamCcamCgaGgtmCa  
 aGa  
 TcamCtaAaaTgcAccAatmClgGacAatmCttmClgmCttmClgmClgGatGcgmc  
 t  
 CEHSP\_F44E5.4/5\_u123\_LNA3 TcaTgaAgcTaaAcaAttmCgaAaaGgaAgaTggTgaAcaAcgGgaAcgTg  
 CEHSP\_F44E5.4/5\_u380\_LNA3 AagTatAacmCttmCcaAcaGggGtcmCgtmCcaGaamCaaAtcAagTccGaaTt  
 CEHSP\_F52E1.7\_u175\_LNA3 TttAacmCatGgcmCgcAgaTtcTtcGatGacGtcGacTttGatmCgcmCacAt  
 CEHSP\_F52E1.7\_u448\_LNA3 GcgTcgAaaAgaTctmCccTgaAgtmClgmCatTgamClgGccTtgAtaTtaTg  
 CEHSP\_F54D5.8\_u252\_LNA3 AcaTagTctTcgTcaTcaAggAtaAgcmCacAccmCgaAatTcaAgcGagAg  
 CEHSP\_F54D5.8\_u318\_LNA3 TcgmcCcaAcamCtcGgamCacGtgmCcaAaaTgaAtaTcaTctmCaaAtcGaaTg  
 CEHUS\_H26D21.1\_u117\_LNA3 GtcGaaGttAgaAatmCcaGaaGccGatAttGttTctmCatmCaaAttmCcaAt  
 CEHUS\_H26D21.1\_u478\_LNA3 ActActmCgtGgaAgaTccAatAaaGttGttTcaAcgmCgamCaaAtcGatTc  
 CEMRE\_ZC302.1\_u169\_LNA3 GgcAgtGaaGatGaaGtgGcaAatTctGatGaaGaaAtgGgaAgcAgtAt  
 CEMRE\_ZC302.1\_u292\_LNA3 TtgTcaAcgAccAgaAgcAaaAatTatGggAatmCgcGatAaaAttmCaaGg  
 CEMTL\_T08G5.10\_d127\_LNA3 GatGcaAgtGtgmCcaActGcgAatGtgmCtcAggmClgmCtcAttAatTtgAa  
 CEMTL\_T08G5.10\_u45\_LNA3 GacGatAtgTtcGatTtcmCcaGgaGagGacGgtGatGatGtgTcaGacTt  
 CENAP\_D2096.8\_u356\_LNA3 GacGatAtgTtcGatTtcmCcaGgaGagGacGgtGatGatGtgTcaGacTt  
 CENAP\_D2096.8\_u70\_LNA3 GagGtcGtcGtaAtcmCacAagGctmCcaAgaAagmCaaGtgmCtcGacAttTc  
 CEPAL\_F56D12.5\_u241\_LNA3 GatActTttGgcAagmCtcGttmCcaAtcAagAagGagGtcAtcmCcaGatmCg  
 CEPAL\_F56D12.5\_u301\_LNA3 GatGagGagGgamCacAccGagmCtcTaaAtcmCacAttmCcaAtamCagTtcAa  
 CEPDI\_C07A12.4\_u28\_LNA3 mCttAtgTccGaaGatAtcmCcaGagGatTggGacAagAacmCcaGtcAagAt  
 CEPDI\_C07A12.4\_u433\_LNA3 TacmCccAgtmCgamCtaTgaTggAgamCagAaamCctmCgaGaaGttmCgaAg  
 aAt  
 CEPDI\_C14B1.1\_u119\_LNA3 mCtcGtcGccTccAacTtcAacGaaAttGccmCttGatGaaAccAagActGt  
 CEPDI\_C14B1.1\_u358\_LNA3 TtcTatTgtTtaTtcmCttGccmCaaTagTgtAttTgtAttTatTctTtcTc  
 CEPGK\_T03F1.3\_d9\_LNA3 mCaaAtcmCatmCtcmCcaGtgGatTtcGtcAttGctGacAagTtcGccGagGa  
 CEPGK\_T03F1.3\_u424\_LNA3 GttTctGatTcgAcamCttTatGgamCcaTctmCaaGttmClgmCgaGttTctTt  
 CEPON\_E01A2.7\_u223\_LNA3 GggAaamCaaAtgAttGttGgtAcaGtaGccmCgcmCctGctAttmCacTgtGa  
 CEPON\_E01A2.7\_u79\_LNA3 mCgaGcamCatmCatmCcaAtcGttmCctGttmCaamCaaGgcmCttmCtaAtcGtt  
 Ag  
 CEPPGB\_F13D12.6\_u44\_LNA3 TgaTgaGagmCccAgtAacmCaaTtaTttGaamCcgTcaGgaTgtGcgTaaGg  
 CEPPGB\_F13D12.6\_u440\_LNA3 mCgtmCtaAtcGaaGaaGggGatmCgtGggmCaaTcaTaamCtaAttAacmCttm  
 Ca  
 CEPPS\_T14G10.1\_d2\_LNA3 mCaaTggmCtcmCagGtcTttmClgmCtcTtcAtaTacTtcmCatTccGagTtgmCt  
 CEPPS\_T14G10.1\_u240\_LNA3 GttmCtcTtgGagmCtgAagTtgTcgcmCgtGctmCgtGtgAttmCtcActTctmCt  
 CEPRDX\_R07E5.2\_u405\_LNA3 TcgcmCtamCcaGcaAggAatActTcaAcaAggTcaAcaAgtGatmCacAcaGa  
 CEPRDX\_R07E5.2\_u42\_LNA3 AagGaaAttGtaActmCgcmCcaAgaGctmCtcmCcaGgtGtcmCgtGgamCatAt  
 CEPYC\_D2023.2\_u256\_LNA3



CEPYC\_D2023.2\_u427\_LNA3  
 CERAD\_F10G7.4\_u169\_LNA3  
 CERAD\_F10G7.4\_u267\_LNA3  
 CERAD\_F32A11.2\_u250\_LNA3  
 CERAD\_F32A11.2\_u380\_LNA3  
 CERAD\_T04H1.4\_u274\_LNA3  
 CERAD\_T04H1.4\_u375\_LNA3  
 CERAD\_W06D4.6\_u325\_LNA3  
 CERAD\_W06D4.6\_u34\_LNA3  
 CERAD\_Y116A8C.13\_u289\_LNA3  
 CERAD\_Y116A8C.13\_u59\_LNA3  
 CERAD\_Y39A1A.23\_u221\_LNA3  
 CERAD\_Y39A1A.23\_u276\_LNA3  
 CERAD\_Y41C4A.14\_u509\_LNA3  
 CERAD\_Y41C4A.14\_u731\_LNA3  
 CERAD\_Y43C5A.6\_u131\_LNA3  
 CERAD\_Y43C5A.6\_u429\_LNA3  
 CERFC\_F31E3.3\_u128\_LNA3  
 CERFC\_F31E3.3\_u55\_LNA3  
 CERPL\_K11H12.2\_d1\_LNA3  
 CERPL\_K11H12.2\_u172\_LNA3  
 CERT\_F36A4.7\_u1396\_LNA3  
 CERT\_F36A4.7\_u2302\_LNA3  
 CERT\_F36A4.7\_u289\_LNA3  
 CERT\_F36A4.7\_u2919\_LNA3  
 CERT\_F36A4.7\_u4269\_LNA3  
 CERT\_F36A4.7\_u5485\_LNA3  
 CESLC\_F52F12.1a\_u249\_LNA3  
 CESLC\_F52F12.1a\_u76\_LNA3  
 CESLC\_K11G9.5\_u400\_LNA3  
 CESLC\_K11G9.5\_u462\_LNA3  
 CESLC\_Y32F6B.1\_u179\_LNA3  
 CESLC\_Y32F6B.1\_u280\_LNA3  
 CESLC\_Y37A1C.1a\_u104\_LNA3  
 CESLC\_Y37A1C.1a\_u404\_LNA3  
 CESLC\_Y70G10A.3\_u383\_LNA3  
 CESLC\_Y70G10A.3\_u46\_LNA3  
 CESOD\_C15F1.7\_u435\_LNA3  
 CESOD\_C15F1.7\_u9\_LNA3  
 CESOD\_F10D11.1\_u326\_LNA3

TtgActGgaTtgGagAttGcgGaaGaaGttGatGttGaaAtcGagAgtGg  
 GccAagTctmCaaGcaAtaAgtGttGatmCaaTcaGagmCcaTacGgaGagAt  
 AtaTtgAgamCttmCggGacAagmCggActTctmCatmCtgTcamCagmCaamCtg  
 mCc  
 GatmCcgmmCagAgaAtcGagTatTtcmCtcTcgAgamCccAtgGatAtcAacTg  
 TccGttAagAagmCtcActGgaAaaAcamCacGgcTcgAacGaaAttGgaAt  
 AatTtgGatGagAgcAaaGtgGaaGgaAtgGctAtcGttTtgGcaGatAt  
 GtgmCtgGtcAaaAaaTgcTtgmCttmCgtTgcTtaTtcGcaTtgmCacTcgmCa  
 mCttmCgaGaamCtcTtcAagTtgGaaTcaAcaGtgGcaTcgGatAcamCatGa  
 GtgmCctTctGaaGccGaaGaaAacGacGatTagTtaAatGttTccAagTt  
 GatAaaAtcGatAgcGacGacGatGagGaaGccGatGatGagGagmCtcGa  
 GcaGgtGgaTacGgaTgtGgaGctGacTttTgcGttTtaTcaAgaAtcTc  
 TccmCgtAgaAgtAgaAatGctAgaAgaAccTgaAcaAgaAgaTcaAgaAa  
 TgcAagAtgTcaGtaTtgAaamCaaTtcmCtgTagAgamCccmCcgAagAaaAt  
 AgtmCtcGtaTccGggAatGttTcaGccTgtGaaAatGctTgtTgaAgamCg  
 mCttmCaaAacmCgtmCgcTttTaaGgaTacAggAacGtgGcamCgcTtcmCgaG  
 g  
 mCagAttGtamCctTcgAaaAggAaaAggAgaGaaTcgmCgtmCgcAaaAatGg  
 TgaTggmCttTgaTtaTtcGagmCagGagmCaaTgaTgtmCcgAgaGtcGttAt  
 mCaaTgamCgaGaaTatTggAgtAatGggGaaActGgtTgcGacTtgmCgaAa  
 TtgGaaAacAatmCtcmCtcGacTttmCtgmCtcActmCttmCgtGaaActAtcmCa  
 TctTgtTatTttAtTtgTttTggGctTgtTccGaaAatGaaAtgGttGt  
 mCaaTggAtcAccAagmCcaGttmCacAagmCacmCgtGagmCaaAgaGgamCt  
 cAc  
 mCttTgtGatGtgAtgActGcgAagGgamCacTtgAtgGctAttAcgAgamCa  
 GagmCcaGctActmCagAtgAcamCtcAacAcgTtcmCatTatGcaGgaGttTc  
 TacActmCcaTccTcgmCcgAcaTacAatmCcaAcaTctmCcamCgcGgaTtcTc  
 AtgGagAagAtgGttTggAtgGaaTgtGggTtgAgaAtcAgaAtaTgcmCg  
 AacmCggGatAccGtgTcgAacGtcAcaTgaAagAtgGcgAtaTaaTcgTc  
 GagGagAttAaamCgcAtgTcaGtgGctmCatGtcGagTttmCcaGaaGtcTa  
 AgaTatTgcmCtcTacTtaTcaTggGccTgaTggmCttTgtmCtgmCcgGtaTt  
 GaaTctmCaamCcamCttmCtgGaamCccmCatAcamCcaAtgGatAgaAgamC  
 ggAg  
 GttGttmCttTttTccGtgAtcTttTcaTgtTtaTgtmCtgAacGtgGcaGg  
 GacTcgTtgGtgTctTgcTagGatGtcTtgGgtTcaTtcmCtcAatmCgtTg  
 GtamCtgGgcTcgAggGctGaaActAatmCgaAgaAgaAacTccAgaAgaTa  
 GgaTcaTgcTctGttTacGacActGatGagTtaAgaGtcAgamCtgmCacGt  
 mCgaTggTtcTtcTcgTctAtcAtaTcgGggTagTtgmCcgAagTgtTgaAa  
 mCaaAtcGaamCtgGtaTaaAggAggAccGacGgaGacGaaTttGaamCgaGa  
 AttmCgaTcaAagAacTctGgcTctmCggmCgtTaamCtgGacAttTgtTcgTc  
 mCtcmCccGagmCagGcgAttAttmCacGctAgtTatGctmCaaAtgTgaTctGt  
 mCcgGtamCtaTctGgaTcamCacAgaAgtmCcgAaaAtgAccAggmCagTtaTt  
 mCccAgtGacTacmCtgAatmCgcGtcTctGaaTctmCcamCacAatTccTacTa  
 GgaGttGctmCacmCgcAatTaaGagmCgamCttmCggAtcTctGgaTaaTctTc

CESOD\_F10D11.1\_u477\_LNA3   AaaTtgAggAaaAgcTtcAcgAggmCggTctmCcaAagGaaAcgTcaAagAa  
                                   mCaaTcgTacmCaiGaaAgaAgtTggAagmCcamCgtGcaAgaGaaGaaAtcmC  
 CESULT\_EEED8.2\_u316\_LNA3   a  
 CESULT\_EEED8.2\_u82\_LNA3    AagAagAttmCctGacmCagAgaGacTcamCgtGctTacmCcaAgaAgcAtcTa  
 CESULT\_Y113G7A.11\_u252\_LNA3  
 3                               AgcAttGgtGgaAatAcgAaaTggmCatGggAagAgaAacmCccTctmCaaTt  
 CESULT\_Y113G7A.11\_u96\_LNA3   mCtgGttAcgGtaGtgTatGgtmCccTgtmCctmCtcAgaAtgmCaaAlaTgtmCg  
 CESULT\_Y67A10A.4\_u108\_LNA3   TctAcgTcgAtgGaaAagmCcgAttTaamCaaTcaAagmCcaAcaAcgmCagTt  
 CESULT\_Y67A10A.4\_u327\_LNA3   GgaAagGtgmCcaAaaAgtTgamCagmCaaTtgGagGatmCttAttmCatTgcmCa  
 CETOPO\_K12D12.1\_u398\_LNA3   AgaTgaTgaTgaAgtTccTgcAaaGaaGccTgcTccAgcGaaGaaAgcTg  
 CETOPO\_K12D12.1\_u449\_LNA3   AaaAccTcgTatTggAaaAggAgcTgcGaaAgcGgaAgtTatmCgaTttGt  
 CETOPO\_M01E5.5b\_u256\_LNA3   GagAagGccmCagAagAagTacGacAgamCtgAagGagmCagTtgAaaAagTt  
 CETOPO\_M01E5.5b\_u429\_LNA3   TtcTgtmCatAcaAtcGtgmCtaAtcGgcAggTtgmCgaTccTttGtaAccAt  
 CEUbi\_F25B5.4\_u186\_LNA3    AagmCttmCggAcamCcaTtgAgaAtgTcaAagmCcaAaaTccAggAtaAggAg  
 CEUbi\_F25B5.4\_u2\_LNA3       AatmCgaAccmCatmCaaTtcActmCgtTatTccTccTcgAtcTccGttmCaaGt  
 CEUbi\_F29B9.6\_u145\_LNA3    mCtgAacmCatmCcaAatAttGaaGatmCcaGctmCagGctGaaGccTatmCagAt  
 CEUbi\_F29B9.6\_u230\_LNA3    mCgtGtgmCttAtcTctTctGgaTgaAaamCaaGgaTtgGaaGccGtcAatmCt  
 CEUbi\_M7.1\_u239\_LNA3       mCggAagmCatmCtgmCctTgamCatTctmCcgTtcGcaGtgGtcGccGgcTctG  
 CEUbi\_M7.1\_u53\_LNA3        AaaGtamCgcTatGtgAggAggmCtaAcamCcaTtcAtaTaaGaamCgcAgcmCa  
 CEUGT\_F39G3.1\_u40\_LNA3      TgtTgcmCgtAgaAgaGagActAaaActAagAacGatTgaTtgAagGtcTg  
 CEUGT\_F39G3.1\_u466\_LNA3    TacAatTctTtgmCagGaaGcaAtaTccGccGgaGtcmCccmCttAtcActAt  
 CEUGT\_M88.1\_u480\_LNA3      mCtcAcgGagGttAtaAttmCtaTgcAggAggmCaaTttmCtgmCtgGagTtcmCa  
 CEUGT\_M88.1\_u72\_LNA3       AccGttTcaTgaGagmCtgTaaTcaGgtGttGttTctGtaAaaAgtGtgAa  
 YAL009W\_u145\_LNA3          GtgGatGtgAaaTtaGtcmCtcAacmCccAgaGcaTttAgtGcaGagAttAg  
 YAL009W\_u341\_LNA3          GcaGttTaaTgtGaaGctAgtTaaAgtAcaGtcTacGtgGgamCgaGaaAt  
 YAL059W\_u262\_LNA3          AttGccAagTccAttTctmCgtGccAagTacAttmCaaAatAcaAgaAagGc  
 YAL059W\_u51\_LNA3          AgamCtcmCtamCaaAtaGatTcgGtgTccTgcmCagAcgAtgTtgAagAatAg  
 YER109C\_u109\_LNA3          TtgAagTttGggAatAttGgtAtgGttGaaGacmCaaGgamCcgGatTacGa  
 YER109C\_u436\_LNA3          GagGcgmCaaGtaGgcAatGatTcaAgaAgtAgtAaaGgcAatmCgtAacAc  
 YHR152W\_u128\_LNA3          TgaGcamCaaAgtTaaGatGttmCggAaaGaaAaaGaaAgtmCaaTccTatGa  
                                   mCaaGtgAccAatmCagmCacGcamCggmCttmCcaTccTcaAgamCtgAtaTta  
                                   mCc  
 YHR152W\_u510\_LNA3          AttAaaTgcGcaGatGagGacGgaAcgAatAtcGgaGaaActGatAatAt  
 YKL130C\_u211\_LNA3          GatGgtAagmCtgAgcGccTtgGacGaaGaaTttGatGttGtcGctActAa  
 YKL130C\_u85\_LNA3            TacGtcAcgmCaaGgamCagAgcTttGacGacGaaAtaTcamCttGgaGgaTt  
 YKL178C\_u199\_LNA3          TctmCccTgtGtaGgtAcamCcaAtaTcamCaaGcgmCatTtcTatGtcGacTa  
 YKL178C\_u367\_LNA3          TgcTaamCacmCagTttAgamCcaTggAaaTccmCacmCgcAaaTatAagmCaa  
                                   Tg  
 YLR443W\_u179\_LNA3          GcaGgamCatAagAttmCcgGtcAagmCaamCgamCagTgaAgaAagTatGcaA  
                                   a  
 YLR443W\_u86\_LNA3            mCcgTctAgtGaaAgcGggAtgGctAaaTtgGgaAaamCgamCaaGatGttAt  
 YOR092W\_u251\_LNA3          GatGctTcaAtaTccTttGatGgtmCgtTagTttAccAttTttGgtGtcTt  
 YOR092W\_u82\_LNA3            AgtmCatTtgAgtTatGtgAagAccGttGgtGggAaaGaaGagAtcAggTg  
 YPL263C\_u132\_LNA3          GtcTtgGctAccAcamCccAaaAccGttmCgaAacTttAagAgcAttmCtamCt

Example 54. Evaluation of different LNA substituted oligonucleotides as probes for fluorescence in situ hybridization (FISH) on metaphase chromosomes and interphase nuclei.

5 Locked Nucleic Acids (LNA) constitute a novel class of DNA analogues that have an exceptionally high affinity towards complementary DNA and RNA. Using human classical satellite-2 repeat sequence clusters as targets, we demonstrate that LNA/DNA mixmer oligonucleotides are excellent probes for FISH combining high binding affinity with short hybridization time and even with the ability to hybridize without prior thermal denaturation of  
10 the template. The development of molecular probes and image analysis has made fluorescence in situ hybridization (FISH) a powerful investigative tool. Although FISH has proved to be a useful technique in many areas, it is a fairly time-consuming procedure with limitations in sensitivity. Probes with higher DNA affinity may potentially reduce the time needed for hybridization and the sensitivity of the technique. Thus, improvement in hybridization  
15 characteristics has been reported for the DNA mimic peptide nucleic acid (PNA). This example describes the development of LNA substituted oligonucleotides as probes for fluorescence in situ hybridization on metaphase chromosomes and interphase nuclei. In each experiment a different LNA substituted oligonucleotide of the same 23-bp human satellite-2 repeat sequence (attccattcgattccattcgatc) have been used, cf. Jeanpierre, M. (1994). Human satellites 2 and 3.  
20 *Annals of Genetics* 37, 163-171. Oligomers with various LNA content, different labels, and hybridization conditions have been used and compared with each other and the optimal conditions have been determined for an efficient LNA-FISH protocol.

#### A. MATERIALS AND METHODS

##### A1. Chromosome Preparations

25 Chromosome preparations were made by standard methods from peripheral lymphocyte cultures of normal males. Slides were prepared 1-4 days prior to an experiment and treated with RNase (10µg/µl) at 37°C for one hour before hybridization.

##### A2. Probe preparation

The 23bp human satellite-2 repeat sequence, attccattcgattccattcgatc, was used to prepare the  
30 LNA/DNA mixmers with different content and sequence order of LNA modifications (Table 19). All mixmers were labeled in the 5' end with either Cy3 or biotin. Biotin amidite was purchased from Applied Biosystems and Cy3 amidite was purchased from Amersham Bioscience. A DNA oligonucleotide of the same sequence without any LNA modifications was used as a control in each experiment.

### A3. Fluorescence in situ hybridization

FISH was carried out as described by Silahtaroglu AN, Hacıhanefioglu S, Guven GS, Cenani A, Wirth J, Tommerup N, Tumer Z. (1998) Not para-, not peri-, but centric inversion of chromosome 12. *J Med Genet.* 35(8):682-4. (1); with the following modifications including:

5 The amounts of probe were 6.4, 10, 13.4 and 20 pmoles. Denaturation of the target DNA and the probe were performed at 75°C for 5 minutes either separately using 70% formamide or simultaneously under the coverslip in the presence of the hybridization mixture containing 50% formamide. In addition the effect of denaturation was also tested. Two alternative hybridization mixtures were used: 50% formamide/2xSSC (pH 7.0) /10% dextran sulphate or 2xSSC (pH 7.0) 10 /10% dextran sulphate. Hybridization times included 30 min, 1 hr, 2 hrs, 3 hrs and overnight. Hybridization temperatures included: 37°C, 55°C, 60°C and 72°C. Post washing was either as for standard FISH (1), or with 50% formamide/2xSSC at 60°C, or without formamide. Hybridization signals with biotin labeled LNA substituted oligonucleotide probes were visualized indirectly using two layers of fluorescein-labeled avidin (Vector Labs) linked by a 15 biotinylated anti-avidin molecule, which amplified the signal 8-64 times. The hybridization of Cy3 labeled molecules however, was visualized directly after a short washing procedure. Slides were mounted in Vectashield (Vector Laboratories) containing 4'-6'-diamidino-2-phenylindole (DAPI). The whole procedure was carried out in the dark. The signals were visualized using a Leica DMRB epifluorescence microscope equipped with a SenSys charge- 20 coupled device camera (Photometrics, Tucson, AZ), and IPLAB Spectrum Quips FISH software (Applied Imaging international Ltd, Newcastle, UK) within two days after hybridization.

### B. RESULTS AND DISCUSSION

Satellite-II DNA, composed of multiple repeats of a 23bp and a 26bp sequence, is especially 25 concentrated in the large heterochromatic regions of human chromosomes 1 and 16, but is also found in the heterochromatic regions of chromosomes 9, Y, 15 and in other minor sites like the short arms/satellites of the acrocentric chromosomes and some centromeric regions. Classical satellite DNA can be visualised by FISH with traditional genomic and DNA oligonucleotide probes (see Kokalj-Vokac N, Alemeida A, Gerbault-Seureau M, Malfoy B, Dutrillaux B. 30 (1993) Two-color FISH characterization of i(1q) and der(1;16) in human breast cancer cells. *Genes Chromosomes Cancer.* 7, 8-14; and Tagarro I, Fernandez-Peralta AM, Gonzales-Aguilera JJ. (1994) Chromosomal localization of human satellites 2 and 3 by a FISH method using oligonucleotides as probes. *Hum Genet.* 93(4):383-8). Due to this and the presence of distinct major and minor sites of satellite-2 DNA in the genome, we used the 23-bp satellite-2 35 repeat sequence, attccattcgattccattcgatc, as a convenient model to test the efficiency of various

DNA/LNA mixmers for FISH analysis and the effect of different experimental conditions by recording the number, location and strength of signals on each metaphase. To compare the efficiency of mixmers with different LNA content (Table 19) and to optimize the LNA-FISH protocol, different conditions were tried at each step of a standard FISH protocol as described in

5 Materials and Methods. All LNA substituted oligonucleotides (LNA/DNA mixmer oligonucleotides) for human satellite-2 sequence gave very prominent signals when used as FISH probes. In general, the signal on chromosome 1 was always stronger and appeared earlier, followed by signals on chromosomes 16, 9, Y, 15, other acrocentric chromosomes and the centromeric regions of other chromosomes, respectively (Figure 51). In general, biotin labeled 10 mixmers gave stronger signals with a higher background, whereas Cy3-labeled molecules gave a significantly lower background.

#### **B1. Effect of LNA content of the LNA substituted oligonucleotides (LNA/DNA mixmers)**

The LNA-2 molecule which had every other nucleotide modified as LNA.

(aTtCcAtTcGaTtCcAtTcGaTc) gave the best results in all the experiment performed. The

15 LNA-3 molecule, with every third oligonucleotide modified as LNA,

(aTtCcAtTcGaTtCcAtTcGaTc) also gave hybridization signals, but with less efficiency than the LNA-2 probes. Preferably, an LNA-2 oligonucleotide molecule has an LNA unit at every other nucleotide position in the sequence and an LNA-3 oligonucleotide molecule has an LNA unit at every third position of the sequence. However, minor deviations, e.g. in one position or less 20 than 5-10 percent of the nucleotide positions in the sequence may still provide the general features of an LNA-2 or an LNA-3 molecule.

The Dispersed LNA (aTtccatTcgaTtccAttcgaTc), which had 5 dispersed LNA modifications, was less efficient in short term hybridization, but gave signals on both chromosomes 1 and 16 after overnight hybridizations. LNA/DNA mixmers with 3 LNA Blocks

25 (aTTCcattcgATTccattcGATc) was comparably inferior as a FISH-probe.

#### **B2. Effect of amount of the LNA/DNA mixmers**

The initial experiments performed with 20 pmol of LNA/DNA mixmer resulted in bright and large signals, but with an extremely high background. Thus, lower concentrations were tested (13.4 pmol, 10 pmol and 6.4 pmol). The concentration giving the optimal signal to noise ratio

30 was found to be 6.4 pmol.

#### **B3. Effect of denaturation**

The signals on the major sites of hybridization (1q, 16q) were equally bright after both types of denaturation. However, smaller and weaker signals were observed on the minor sites with the simultaneous denaturation protocol. To check the potential "strand invasion" property of LNA,

35 some of the experiments were performed without a denaturation step. As expected, no signals

were obtained by the control DNA oligonucleotide probe. In contrast, hybridization signals on chromosomes 1 and 16 were observed after overnight hybridization with LNA probes, with LNA-2 mixmer giving the best signals. Compared to the signals obtained in experiments involving a denaturation step, the signals were smaller, but prominent and without any background.

#### **B4. Effect of hybridization time, temperature and post-hybridization washes**

Although signals could be observed after only 30 min of hybridization, the optimal hybridization time and temperature for LNA-2, which gave the best signals, was 1 hr at 37°C. A 3x5 min wash with 0.1xSSC/60°C and 4xSSC/0.05%Tween/37°C, respectively, followed by a 5 min PBS wash was found to be sufficient for washing the slides after hybridization with DNA-LNA mixmers. There was no specific difference between a wash with 50% formamide at 42°C or 60°C.

The signals faded away in most of the slides within two days. When hybridized with directly labeled LNA, the whole slide was stained with Cy3 after three days. Thus, slides had to be analyzed within 48 hours after hybridization.

#### **C. CONCLUSION**

The experiments have demonstrated, that LNA substituted oligonucleotides are very efficient FISH probes. LNA substituted oligonucleotide probes gave strong signals after only 1 hr of hybridization, and it was possible to omit the use of formamide both from the denaturation and from the post hybridization washing steps and still obtain a very good signal to noise ratio. The ability of LNA to hybridize without prior denaturation could be due to a strand invasion property of LNA and this warrants further investigation with other LNA probes and at different conditions. Based on the combined results of these experiments, the optimal LNA-FISH procedure was defined as follows: 6.4 pmoles of Cy-3 labeled LNA-2 probe was denatured together with the target at 75°C for 5 minutes, and hybridized for one hour then followed by a short post wash without any formamide (3x 5 minutes 0.1xSSC at 60°C; 2x 5 minutes 4xSSC/0.05%Tween at 37°C; 5 minute PBS). The FISH experiments indicate that LNA containing probes would be valuable for the detection of a variety of other repetitive elements, such as centromeric  $\alpha$ -repeats and telomeric repeats. In addition, the superior hybridization characteristics of LNA containing oligonucleotides could lead to detection of single base pair differences between repetitive sequences as well as single copy sequences.

**C1.** Figure 51 shows a comparison of different LNA/DNA mixmer oligonucleotides.

Experiment conditions: 6.4 pmoles of Cy3 labeled probe was hybridized for 30 minutes at 37°C, after simultaneous denaturation of the target and the probe at 75°C for 5 minutes. A.

LNA-2 giving signals on chromosomes 1, 16, 9 and 15, B. LNA-3 giving bright signals on chromosomes 1, 16 and 9, C. Dispersed LNA giving signals on chromosomes 1 and 16 only, D. LNA Block giving smaller signals on chromosome 1, E. DNA oligo giving no signals on any of the chromosomes.

5

Table 19. DNA/LNA mixmers for human satellite 2 repeat sequence used in this study.

Name	FISH probe sequences	LNA monomers	Tm*
DNA oligo	attccattcgattccattcgatc	0	60
Dispersed LNA	aTtccatTcgaTtccAttcgaTc	5	71
LNA-3	aTtcCatTcgAtTccAttCgaTc	8	77
LNA Blocks	aTTCcattcgATTccattcGATc	9	73
LNA-2	ATtCcAtTcGaTtCcAtTcGaTc	11	84

LNA modifications are depicted in capital letters and \*Tm values for each molecule have been calculated using Exiqon's Tm Prediction program accessible at <http://lna-tm.com/> and as

10 appears from Figure 27 herein.

#### Other Embodiments

From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions.

The foregoing description of the invention is merely illustrative thereof, and it understood that 15 variations and modifications can be effected without departing from the scope or spirit of the invention.

All publications, patent applications, and patents mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication, patent, or patent application was specifically and individually indicated to be incorporated by 20 reference.

**Claims**

1. A nucleic acid that is a non-naturally-occurring nucleic acid with a melting  
5 temperature that is at least 3°C higher than that of the corresponding control nucleic acid  
with 2'-deoxynucleotides; wherein said nucleic acid hybridizes to a first region within a  
first exon of a target nucleic acid and to a second region within a second exon of said  
target nucleic acid that is adjacent to said first exon.
2. A nucleic acid that is a non-naturally-occurring nucleic acid with a melting  
10 temperature that is at least 3°C higher than that of the corresponding control nucleic acid  
with 2'-deoxynucleotides; wherein said nucleic acid hybridizes to a first region within  
an exon of a target nucleic acid and to a second region within an intron of said target  
nucleic acid that is adjacent to said exon.
3. A nucleic acid that is a non-naturally-occurring nucleic acid with a melting  
15 temperature that is at least 3°C higher than that of the corresponding control nucleic acid  
with 2'-deoxynucleotides; wherein said nucleic acid hybridizes to a first region within a  
first intron of a target nucleic acid and to a second region within a second intron of said  
target nucleic acid that is adjacent to said first intron.
4. A nucleic acid that is a non-naturally-occurring nucleic acid with a capture  
20 efficiency that is at least 10% greater than that of a corresponding control nucleic acid  
with 2'-deoxynucleotides at the temperature equal to the melting temperature of said  
nucleic acid; wherein said nucleic acid hybridizes to a first region within a first exon of  
a target nucleic acid and to a second region within a second exon of said target nucleic  
acid that is adjacent to said first exon.
- 25 5. A nucleic acid that is a non-naturally-occurring nucleic acid with a capture  
efficiency that is at least 10% greater than that of a corresponding control nucleic acid  
with 2'-deoxynucleotides at the temperature equal to the melting temperature of said  
nucleic acid; wherein said nucleic acid hybridizes to a first region within an exon of a  
target nucleic acid and to a second region within an intron of said target nucleic acid that  
30 is adjacent to said exon.
6. A nucleic acid that is a non-naturally-occurring nucleic acid with a capture  
efficiency that is at least 10% greater than that of a corresponding control nucleic acid  
with 2'-deoxynucleotides at the temperature equal to the melting temperature of said  
35 nucleic acid; wherein said nucleic acid hybridizes to a first region within a first intron of



a target nucleic acid and to a second region within a second intron of said target nucleic acid that is adjacent to said first intron.

7. A nucleic acid that is an LNA and that hybridizes to a first region within a first exon of a target nucleic acid and to a second region within a second exon of said target nucleic acid that is adjacent to said first exon.

8. A nucleic acid that is an LNA and that hybridizes to a first region within an exon of a target nucleic acid and to a second region within an intron of said target nucleic acid that is adjacent to said exon.

9. A nucleic acid that is an LNA and that hybridizes to a first region within a first intron of a target nucleic acid and to a second region within a second intron of said target nucleic acid that is adjacent to said first intron.

10. The nucleic acid of any one of claims 1-9, wherein the length of said first region and the length of said second region are between 3 and 50 nucleotides, inclusive.

11. The nucleic acid of claim 10, wherein the length of said first region and the length of said second region are between 10 and 40 nucleotides, inclusive.

12. The nucleic acid of claim 11, wherein the length of said first region and the length of said second region are between 20 and 30 nucleotides, inclusive.

13. The nucleic acid of any one of claims 1-9, wherein said first region and said second region are the same length.

14. The nucleic acid of any one of claims 1-9, wherein said first region and said second region are a different length.

15. A population of nucleic acids that includes a nucleic acid of any one of claims 1-14.

16. A population of nucleic acids that includes a non-naturally-occurring nucleic acid with a melting temperature that is at least 3°C higher than that of the corresponding control nucleic acid with 2'-deoxynucleotides; wherein said nucleic acid hybridizes to only one exon or to only one intron of a target nucleic acid.

17. A population of nucleic acids that includes a non-naturally-occurring nucleic acid with a capture efficiency that is at least 10% greater than that of a corresponding control nucleic acid with 2'-deoxynucleotides at the temperature equal to the melting temperature of said nucleic acid; wherein said nucleic acid hybridizes to only one exon or to only one intron of a target nucleic acid.

18. A population of nucleic acids that includes a nucleic acid that is an LNA and that hybridizes to only one exon or to only one intron of a target nucleic acid.

19. A nucleic acid of any one of claims 1-9 that is a non-naturally occurring nucleic acid with a melting temperature that is at least 3°C higher than that of the corresponding control nucleic acid with 2'-deoxynucleotides; wherein said nucleic acid hybridizes to only one exon or to only one intron of a target nucleic acid.
20. A nucleic acid of any one of claims 1-9 that is a non-naturally-occurring nucleic acid with a capture efficiency that is at least 10% greater than that of a corresponding control nucleic acid with 2'-deoxynucleotides at the temperature equal to the melting temperature of said nucleic acid; wherein said nucleic acid hybridizes to only one exon or to only one intron of a target nucleic acid.
21. A nucleic acid of any one of claims 1-9 that comprises an LNA and that hybridizes to only one exon or to only one intron of a target nucleic acid.
22. The population of any one of claims 15-18, wherein said nucleic acid is between 15 and 150 nucleotides in length, inclusive.
23. The population of claim 22, wherein said nucleic acid is between 5 and 100 nucleotides in length, inclusive.
24. The population of claim 23, wherein said nucleic acid is between 20 and 80 nucleotides in length, inclusive.
25. The population of claim 24, wherein said nucleic acid is between 30 and 60 nucleotides in length, inclusive.
26. The population of claim 25, wherein said nucleic acid is 40 nucleotides in length.
27. The population of claim 25, wherein said nucleic acid is 50 nucleotides in length.
28. The nucleic acid of any one of claims 1-9 and 19-21 wherein said nucleic acid is between 8 and 70 nucleotides in length.
29. The nucleic acid of any one of claims 1-9 and 19-21 wherein said nucleic acid is between 9 and 50 nucleotides in length.
30. The nucleic acid of any one of claims 1-9 and 19-21 wherein said nucleic acid is between 12 and 40 nucleotides in length.
31. The nucleic acid of any one of claims 1-9 and 19-21 wherein said nucleic acid is between 15 and 35 nucleotides in length.
32. The population of any one of claims 15-18, wherein at least 5% of the nucleotides in said nucleic acid are LNA units.

33. The population of claim 32, wherein at least 10% of the nucleotides in said nucleic acid are LNA units.
34. The population of claim 32, wherein at least 20% of the nucleotides in said nucleic acid are LNA units.
- 5 35. The population of any one of claims 15-18, wherein every second nucleotide in said nucleic acid is an LNA unit.
36. The population of any one of claims 15-18, wherein every third nucleotide in said nucleic acid is an LNA unit.
37. The population of any one of claims 15-18, wherein every fourth nucleotide in  
10 said nucleic acid is an LNA unit.
38. The population of any one of claims 15-18, wherein every fifth nucleotide in said nucleic acid is an LNA unit.
39. The population of any one of claims 15-18, wherein every sixth nucleotide in said nucleic acid is an LNA unit.
- 15 40. The population of any one of claims 15-18, wherein (i) every second and every third nucleotide, (ii) every second and every fourth nucleotide, (iii) every second and every fifth nucleotide, (iv) every second and every sixth nucleotide, (v) every third and every fourth nucleotide, (vi) every third and every fifth nucleotide, (vii) every third and every sixth nucleotide, (viii) every fourth and  
20 every fifth nucleotide, (ix) every fourth and every sixth nucleotide, and/or (x) every fifth and every sixth nucleotide in said nucleic acid is an LNA unit.
41. The population of claim 40, wherein every second, every third, and every fourth nucleotide in said nucleic acid is an LNA unit.
42. The population of any one of claims 15-18, wherein said nucleic acid comprises  
25 two or more contiguous LNA units.
43. The population of claim 42, wherein said nucleic acid comprises at least 4 contiguous LNA units.
44. The population of claim 42, wherein said nucleic acid comprises at least 5 contiguous LNA units.
- 30 45. The population of claim 42, wherein the number of contiguous LNA units is between 5 and 20% of the total length of said nucleic acid.
46. The population of claim 45, wherein the number of contiguous LNA units is between 10 and 15% of the total length of said nucleic acid.
47. The population of claim 42, wherein at least one LNA unit in said nucleic acid  
35 hybridizes to a first region within a first exon of a target nucleic acid and at least

one LNA unit in said nucleic acid hybridizes to a second region within a second exon of said target nucleic acid that is adjacent to said first exon.

48. The population of claim 47, wherein at least two LNA units in said nucleic acid hybridize to a first region within a first exon of a target nucleic acid and at least two LNA units in said nucleic acid hybridize to a second region within a second exon of said target nucleic acid that is adjacent to said first exon.

49. The population of claim 48, wherein at least three LNA units in said nucleic acid hybridize to a first region within a first exon of a target nucleic acid and at least three LNA units in said nucleic acid hybridize to a second region within a second exon of said target nucleic acid that is adjacent to said first exon.

50. The population of any one of claims 15-18, wherein the 5' terminal nucleotide of said nucleic acid is not an LNA unit.

51. The population of any one of claims 15-18, wherein the 5' terminal nucleotide of said nucleic acid is an LNA unit.

52. The population of any one of claims 15-18, wherein the 3' terminal nucleotide of said nucleic acid is not an LNA unit.

53. The population of any one of claims 15-18, wherein said nucleic acid can distinguish between different nucleic acids that cannot be distinguished using a naturally-occurring control nucleic acid.

54. The population of claim 53, wherein said control nucleic acid consists of only 2'-deoxynucleotides.

55. The population of claim 53, wherein said different nucleic acids are mRNA splice variants.

56. The population of any one of claims 15-18, wherein said nucleic acid comprises one or more universal bases.

57. The population of claim 56, wherein said universal base is located at the 5' or 3' terminus of said nucleic acid.

58. The population of claim 56, wherein one or more universal base are located at the 5' and 3' termini of said nucleic acid.

59. The population of claim 56, wherein all of said nucleic acids of said population have the same number of universal bases.

60. The population of claim 56, wherein said universal base is inosine, pyrene, 3-nitropyrrole, or 5-nitroindole.

61. The population of any one of claims 15-18, wherein said nucleic acid comprises at least one LNA A or LNA T.

62. The population of any one of claims 15-18, wherein each nucleic acid in said population comprises at least one LNA A or LNA T.
63. The population of claim 61, wherein all of the adenine and thymine-containing nucleotides in said LNA are LNA A and LNA T, respectively.
- 5 64. The population of any one of claims 15-18, wherein said nucleic acid comprises a at least one 2,6,-diaminopurine or 2-thio-thymine base.
65. The population of any one of claims 15-18, wherein at least 5% of the nucleic acids in said population are LNA.
66. The population of any one of claims 15-18, wherein at least 10% of the nucleic acid in said population are LNA.
- 10 67. The population of any one of claims 15-18, that includes nucleic acids that together hybridize to at least 10% of the nucleic acids expressed by a particular cell or tissue.
68. The population of claim 67, that includes nucleic acids that together hybridize to at least 50% of the exons of a target nucleic acid.
- 15 69. The population of any one of claims 15-18, wherein said nucleic acid does not form a hairpin that would otherwise inhibit its binding to a target nucleic acid.
70. The population of any one of claims 15-18, wherein opposing nucleotides in a palindrome pair or opposing nucleotides in inverted repeats of said nucleic acid are not both LNA units.
- 20 71. The population of any one of claims 15-18, wherein said nucleic acid forms less than 3 intramolecular base-pairs.
72. The population of any one of claims 15-18, wherein said nucleic acid does not have LNA-5-nitroindole: LNA-5-nitroindole intramolecular base-pairs.
- 25 73. The population of any one of claims 15-18, wherein said nucleic acid has a LNA unit with a 2,6,-diaminopurine, 2-aminopurine, 2-thio-thymine, 2-thio-uracil, inosine, or hypoxanthine base.
74. The population of any one of claims 15-18, wherein said nucleic acid has a 2'O, 4'C-methylene linkage.
- 30 75. The population of any one of claims 15-18, wherein one or more nucleic acids are LNA/DNA, LNA/RNA, or LNA/DNA/RNA chimeras.
76. The population of any one of claims 15-18, wherein said nucleic acid has a melting temperature that is at least 10°C higher than that of the corresponding control nucleic acid with 2'-deoxynucleotides.

77. The population of claim 76, wherein said nucleic acid has a melting temperature that is at least 20°C higher than that of the corresponding control nucleic acid with 2'-deoxynucleotides.

5 78. The population of any one of claims 15-18, wherein said nucleic acid has a capture efficiency that is at least 100% greater than that of the corresponding control nucleic acid with 2'-deoxynucleotides at the temperature equal to the melting temperature of said nucleic acid.

10 79. The population of claim 78, wherein said nucleic acid has a capture efficiency that is at least 400% greater than that of the corresponding control nucleic acid with 2'-deoxynucleotides at the temperature equal to the melting temperature of said nucleic acid.

80. The population of any one of claims 15-18, wherein said nucleic acids are covalently bonded to a solid support.

15 81. The population of claim 80, wherein said nucleic acids are in a predefined arrangement.

82. The population of any one of claims 15-18, comprising at least 10 different nucleic acids.

83. The population of claims 82, comprising at least 100 different nucleic acids.

84. The population of claim 83, comprising at least 500 different nucleic acids.

20 85. The population of claim 84, comprising at least 1,000 different nucleic acids.

86. The population of claim 85, comprising at least 5,000 different nucleic acids.

25 87. A complex of one or more target nucleic acids and one or more nucleic acids or populations of any one of claims 1-9, 15-18 or 19-21 in which one or more target nucleic acids are hybridized to one or more said nucleic acids or populations.

88. The complex of claim 87, wherein at least 10 different target nucleic acids are hybridized.

89. The complex of claim 87, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample.

30 90. The complex of claim 87, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase.

35 91. The complex of claim 87, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 300 nucleotides.

92. The complex of claim 87, wherein the target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 200 nucleotides.
- 5 93. The complex of claim 87, wherein the target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 100 nucleotides.
94. The complex of claim 87, wherein the target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 50 nucleotides.
- 10 95. The complex of claim 87, wherein said target nucleic acids are cRNA molecules amplified from a patient sample.
96. The complex of claim 87, wherein said target nucleic acids are cRNA molecules amplified from a patient sample and fragmented using alkaline hydrolysis.
- 15 97. The complex of claim 87, wherein at least 10 different target nucleic acids are labeled and hybridized.
98. The complex of claim 87, wherein said target nucleic acids are RNA molecules isolated from a patient sample and labeled chemically directly, e.g. using platinum-linked fluorescent dyes.
- 20 99. The complex of claim 87, wherein said target nucleic acids are RNA molecules isolated from a patient sample and labeled chemically directly, e.g. using platinum-linked fluorescent dyes and fragmented using alkaline hydrolysis.
- 25 100. A method for detecting the presence of one or more target nucleic acids in a sample, said method comprising incubating a labeled nucleic acid sample with one or more first nucleic acids of any one of claims 1-9 or 19-21 or one or more populations of first nucleic acids of any one of claims 15-18 under conditions that allow at least one target nucleic acid to hybridize to at least one of said first nucleic acids.
- 30 101. The method of claim 100, wherein hybridization is detected between at least 5 target nucleic acids and said first nucleic acids.
102. The method of claim 100, further comprising identifying one or more hybridized target nucleic acids.
103. The method of claim 100, further comprising determining the amount of one or more hybridized target nucleic acids.
- 35 104. The method of claim 100, wherein one or more target nucleic acids are labeled with a fluorescent group, and wherein the determination of the amount of said

hybridized target nucleic acid involves one or more of the following: (i) adjusting for the varying intensity of an excitation light source used for detection of said hybridization, (ii) adjusting for photobleaching of said fluorescent group, and/or (iii) comparing the fluorescent intensity of said hybridized target nucleic acid(s) to the fluorescent intensity of a different sample of hybridized nucleic acids.

105. The method of claim 100, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample.

106. The method of claim 105, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase.

107. The method of claim 105, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 300 nucleotides.

108. The method of claim 105, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 200 nucleotides.

109. The method of claim 105, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 100 nucleotides.

110. The method of claim 105, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 50 nucleotides.

111. The method of claim 105, wherein said target nucleic acids are cRNA molecules amplified from a patient sample that is optionally fragmented using alkaline hydrolysis.

112. The method of claim 100, further comprising determining the presence or absence of an mRNA splice variant of interest in said sample or determining the presence or absence of a mutation, deletion, and/or duplication of an exon of interest.

113. The method of claim 100, wherein the sample has nucleic acids that are amplified using one or more primers specific for an exon of a target nucleic acid, and wherein said method involves determining the presence or absence of an mRNA splice variant with said exon in said sample.



114. The method of claim 113, wherein one or more of said primers are specific for an exon or exon-exon junction of a pathogen of interest, and said method involves determining the presence or absence of a nucleic acid with said exon in said sample.

5 115. The method of claim 100, wherein said first nucleic acids are covalently bonded to a solid support by reaction of a nucleoside phosphoramidite with an activated solid support, and subsequent reaction of a nucleoside phosphoramidite with an activated nucleotide or nucleic acid bound to said solid support.

10 116. The method of claim 100, further comprising contacting said target nucleic acid with a second nucleic acid or a population of second nucleic acids that binds to a different region of the target molecule than said first nucleic acid.

117. A method for amplifying a target nucleic acid, said method comprising the steps of: (a) incubating one or more first nucleic acids of any one of claims 1-9 or 19-21 or one or more populations of first nucleic acids of any one of claims 15-18 with a target nucleic acid under conditions that allow said first nucleic acid to bind said target nucleic acid; and (b) extending said first nucleic acid with said target nucleic acid as a template.

118. The method of claim 117, further comprising contacting said target nucleic acid with a second nucleic acid or a population of second nucleic acids that binds to a different region of the target molecule than said first nucleic acid.

119. A method for classifying a test nucleic acid sample comprising a target nucleic acid, said method comprising the steps of: (a) incubating a test nucleic acid sample with one or more nucleic acids probes of any one of claims 1-9 or 19-21 or one or more populations of nucleic acids probes of any one of claims 15-18 under conditions that allow at least one of the nucleic acids in said test sample to hybridize to at least one nucleic acid probe; (b) detecting a hybridization pattern of said test nucleic acid sample; and (c) comparing said hybridization pattern to a hybridization pattern of a first nucleic acid standard, whereby said comparison indicates whether or not said test sample has the same classification as said first standard.

120. The method of claim 119, further comprising comparing a hybridization pattern of said test nucleic acid sample to a hybridization pattern of a second standard.

121. The method of claim 119, wherein at least 5 target nucleic acids hybridize to said nucleic acid probes.

122. The method of claim 119, further comprising identifying a hybridized target nucleic acid.
123. The method of claim 119, further comprising determining the amount of said hybridized target nucleic acid.
- 5 124. The method of claim 119, wherein said target nucleic acids are labeled with fluorescent groups.
125. The method of claim 124, wherein said determination comprises scaling for the varying labeling efficiency for the different fluorescent groups used for detection of said hybridization.
- 10 126. The method of claim 124, wherein said determination comprises adjusting for the varying intensity of the excitation light source used for detection of said hybridization.
127. The method of claim 124, wherein said determination comprises adjusting for photobleaching of said fluorescent group.
- 15 128. The method of claim 124, wherein said comparison comprises adjusting for a difference in the amount of said nucleic acid probes used for hybridization to said test sample and said first standard.
129. The method of claim 124, wherein said comparison comprises adjusting for a difference in the buffer used for hybridization to said test sample and said first standard.
- 20 130. The method of claim 129, wherein said difference is a difference in  $Mg^{2+}$  concentration.
131. The method of claim 124, wherein said first nucleic acid standard is labeled with a different fluorescent group.
- 25 132. The method of claim 119, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and optionally fragmented using *E. coli* Uracil-DNA Glycosylase.
133. The method of claim 119, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 300 nucleotides.
- 30 134. The method of claim 119, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 200 nucleotides.

135. The method of claim 119, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 100 nucleotides.
136. The method of claim 119, wherein said target nucleic acids are cDNA molecules reverse transcribed from a patient sample and fragmented using *E. coli* Uracil-DNA Glycosylase to an average size of 50 nucleotides.
137. The method of claim 119, wherein said target nucleic acids are cRNA molecules amplified from a patient sample.
138. The method of claim 119, wherein said target nucleic acids are cRNA molecules amplified from a patient sample and fragmented using alkaline hydrolysis.
139. The method of claim 119, further comprising determining the presence or absence of an mRNA splice variant of interest in said sample.
140. The method of claim 119, further comprising determining the presence or absence of a mutation, deletion, and/or duplication of an exon of interest.
141. The method of claim 140, wherein said mutation, deletion, and/or duplication is indicative of a disease, disorder, or condition.
142. The method of claim 141, wherein said disease is cancer.
143. The method of claim 119, wherein the sample has nucleic acids that are amplified using one or more primers specific for an exon of a target nucleic acid, and wherein said method involves determining the presence or absence of an mRNA splice variant with said exon in said sample.
144. The method of claim 143, wherein one or more of said primers are specific for an exon or exon-exon junction of a pathogen of interest, and said method involves determining the presence or absence of a nucleic acid with said exon in said sample.
145. The method of claim 119, wherein said first nucleic acids are covalently bonded to a solid support by reaction of a nucleoside phosphoramidite with an activated solid support, and subsequent reaction of a nucleoside phosphoramidite with an activated nucleotide or nucleic acid bound to said solid support.
146. A method of selecting a nucleic acid for a population of nucleic acids, said method comprising the steps of: (a) determining the melting temperature of a nucleic acid, determining the ability of said nucleic acid to self-anneal, determining the ability of said nucleic acid to hybridize to one or more exons or introns of a target nucleic acid, and/or determining the ability of said nucleic

acid to hybridize to a non-target nucleic acid, and (b) selecting said nucleic acid for inclusion or exclusion from said population based on the determination in step (a), wherein said nucleic acid is a nucleic acid of any one of claims 1-9 or 19-21 or a nucleic acid that has least one LNA unit and that hybridizes to only one exon or to only one intron of a target nucleic acid.

147. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 for the detection, amplification, or classification of a nucleic acid of interest or a population of nucleic acids of interest.

148. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 for alternative mRNA splice variant detection, expression profiling, comparative genomic hybridization, or real-time PCR.

149. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 as a PCR primer or FISH probe.

150. Use according to claim 149 wherein said nucleic acid or said population of nucleic acids comprise LNA in about every other or about every third position of the nucleotide sequence.

151. Use according to claim 149 wherein said nucleic acid or said population of nucleic acids are used for the detection of a repetitive element.

152. Use according to claim 151 wherein said repetitive element is a centromeric alpha-repeat or a telomeric repeat.

153. Use according to claim 149 wherein said nucleic acid or said population of nucleic acids are used for the detection of a single base pair difference between repetitive sequences or for the detection of single copy sequences.

154. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 as a probe in fluorescent in situ hybridisation of a target DNA without a denaturation step.

155. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 in a homogeneous assay such as quantitative real-time RT-PCR assay, e.g. Taqman real-time RT-PCR, Lightcycler real-time RT-PCR, or real-time NASBA (nucleic acid sequence-based amplification) assay with a Molecular Beacon-based detection.

156. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 in a diagnostic kit based on a

homogeneous assay such as quantitative real-time RT-PCR assay, e.g. Taqman real-time RT-PCR, Lightcycler real-time RT-PCR, or real-time NASBA (nucleic acid sequence-based amplification) assay with a Molecular Beacon-based detection.

- 5 157. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 as an antisense, antigene or ribozyme or double stranded nucleic acid therapeutic agent, or in therapeutic use as modulating and silencing sense nucleic acid agents.
- 10 158. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 as an antisense, antigene or ribozyme or double stranded nucleic acid therapeutic agent, in which the said nucleic acid or population of nucleic acids hybridize to specific splice isoforms or isoforms.
- 15 159. Use of a nucleic acid or a population of nucleic acids of any one of claims 1-9 or 19-21 or of claims 15-18 as an antisense, antigene or ribozyme or double stranded nucleic acid therapeutic agent, in which the said nucleic acid or population of nucleic acids hybridize to a non-coding antisense RNA or RNAs.
- 20 160. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 in a diagnostic kit, based on detection of diagnostic splice isoforms or splice patterns, or diagnostic nucleic acids, such as viral or bacterial mRNAs.
- 25 161. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 in a diagnostic microarray based on detection of mRNA signatures or splice isoform signatures in the concentration range of 10000 ppm to 1000 ppm.
- 30 162. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 in a diagnostic microarray based on detection of mRNA signatures or splice isoform signatures in the concentration range of 1000 ppm to 100 ppm.
163. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 in a diagnostic microarray based on detection of mRNA signatures or splice isoform signatures in the concentration range of 100 ppm to 10 ppm.
164. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 in a diagnostic microarray based on

detection of mRNA signatures or splice isoform signatures in the concentration range of 10 ppm to 1 ppm.

165. Use of a nucleic acid of any one of claims 1-9 or 19-21 or a population of nucleic acids of any one of claims 15-18 in a diagnostic microarray based on detection of mRNA signatures or splice isoform signatures.

166. Use of a non-naturally occurring nucleic acid of any one of claims 1-9 or 19-21 or use of a population of non-naturally occurring nucleic acids of any one of claims 15-18 in a method for generating an array of said non-naturally occurring nucleic acids of chosen lengths within discrete locations of a support material.

167. Use according to claim 166, where said non-naturally occurring nucleic acids are LNA oligonucleotides.

168. Use according to any one of claims 166-167, where said non-naturally occurring nucleic acids are LNA oligonucleotides containing a photochemically active group at the 5' end or the 3' end, separated by a linker from the LNA oligonucleotides.

169. Use according to any one of claims 166-168, wherein said linker is a hexaethylene monomer, dimer, trimer, tetramer, pentamer or hexamer, or a poly-T sequence of 10-50 nucleotides in length, or a poly-C sequence of 10-50 nucleotides in length.

170. Use according to any one of claims 168-169 for generating an array comprising the steps of a) selecting said nucleic acid or population of nucleic acids according to the method of claim 146, b) synthesis of the said nucleic acids using phosphoramidite chemistry and DNA and LNA phosphoramidites; c) purification of the said nucleic acids; d) printing of the said nucleic acids of chosen lengths and sequence onto a polymer surface; and e) coupling of said nucleic acids via excitation of the photochemically active group at the 5' end or the 3' end covalently onto the polymer surface using UV light.

171. The method of claims 166-170 wherein a computer-controlled microarray printing robot delivers the said nucleic acids onto discrete locations on the polymer surface.

172. The method of claims 166-171, wherein the size of each discrete location is between an average size of 10 and 150 microns.

173. The method of claims 166-172, wherein said nucleic acids are printed at a density of at least 50 nucleic acids per square cm.

174. The method of claims 166-172, wherein said nucleic acids are printed at a density of at least 100 nucleic acids per square cm.
175. The method of claims 166-172, wherein said nucleic acids are printed at a density of at least 500 nucleic acids per square cm.
- 5 176. The method of claims 166-172, wherein said nucleic acids are printed at a density of at least 1000 nucleic acids per square cm.
177. The method of claims 166-172, wherein said nucleic acids are printed at a density of at least 2500 nucleic acids per square cm.
- 10 178. The method of claims 166-172, wherein said nucleic acids are printed at a density of at least 4000 nucleic acids per square cm.

# OLIGONUCLEOTIDES USEFUL FOR DETECTING AND ANALYZING NUCLEIC ACIDS OF INTEREST

## Abstract of the Disclosure

5 The invention features improved nucleic acids and methods for expression profiling of mRNAs, identifying and profiling of particular mRNA splice variants, and detecting mutations, deletions, or duplications of particular exons or other splice variants, e.g., alterations associated with a disease such as cancer, in a nucleic acid sample, e.g., a biological sample or a patient sample.

10



Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

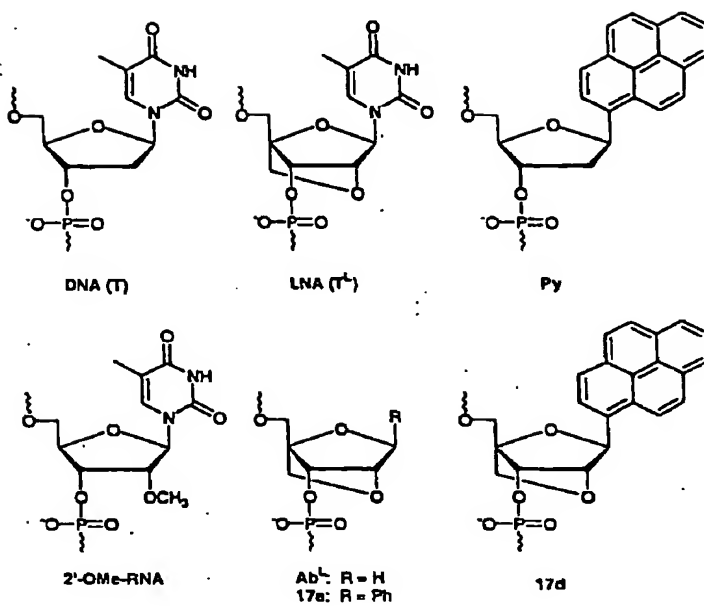


Figure 1

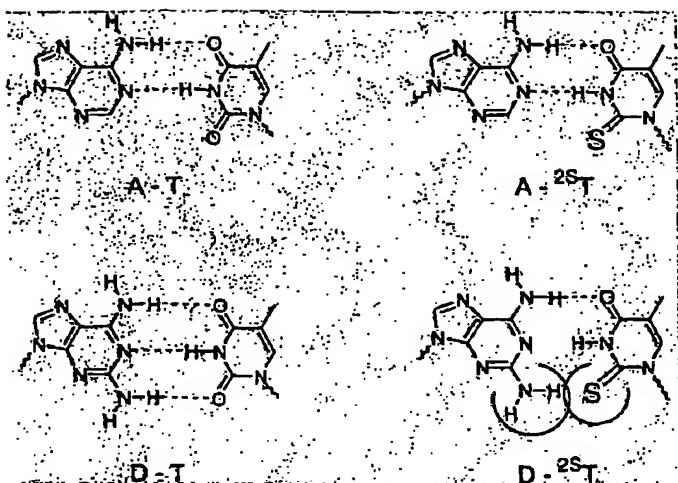


Figure 2

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

19 MAJ 2003

Modtaget

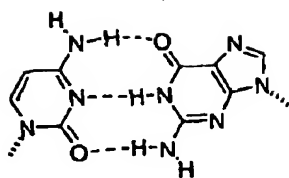
Melting temperatures ( $T_m$ ) of the complementary DNA-DNA and LNA-DNA duplexes.<sup>a</sup>  
Modified monomers (LNA are in CAPITALS): I = inosine; D = 2,6-diaminopurine; X = 2-aminopurine.

Entry	Oligonucleotide structure	$T_m$ ( $\pm 0.5$ °C) of the duplexes with complementary deoxynucleotide			
		3'-ctgtatcc	3'-ctgaaatcc	3'-ctggatcc	3'-ctgeatcc
1	5'-gacatagg	23.8	<10	<10	<10
2	5'-gacttagg	<10	22.6	<10	<10
3	5'-gacgtagg	<10	<10	<10	25.0
4	5'-gacdtagg	23.3	<10	<10	<10
5	5'-gdcddtgg	33.4	<10	<10	17.7
6	5'-gacitagg	<10	<10	<10	20.9
7	5'-gacxtagg	<10	<10	<10	<10
8	5'-GACATAGG	61.6	38.2	43.4	40.6
9	5'-GACTTAGG	28.0	60.7	36.4	23.5
10	5'-GACGTAGG	55.0	32 <sup>b</sup>	41 <sup>b</sup>	70.9
11	5'-GACDTAGG	67.8	42.2	41.4	52.4
12	5'-GDCDTDGG	78.3	55.9	54.7	63.8
13	5'-GACITAGG	53.1	48.2	43.0	59.9
14	5'-GACXTAGG	60.8	45.5	44.0	53.9

<sup>a</sup> The melting temperatures ( $T_m$  values) were obtained as a maxima of the first derivatives of the corresponding melting curves (optical density at 260 nm versus temperature). Concentration of the duplexes: 2.5  $\mu$ M. Buffer: 0.1 M NaCl; 10 mM Na-phosphate (pH 7.0); 1 mM EDTA.

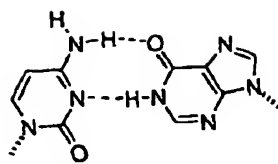
<sup>b</sup> Low cooperativity of transitions (accuracy  $\pm 1$  °C).

Figure 3



C

G



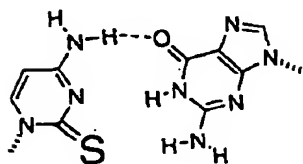
C

I

Patent- og  
Varemærkestyrelsen

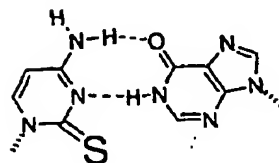
19 MAJ 2003

Modtaget



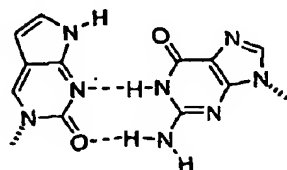
2sC

G



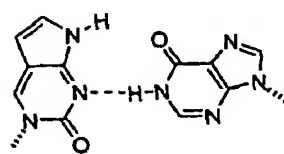
2sC

I



PyrroloPyr

G



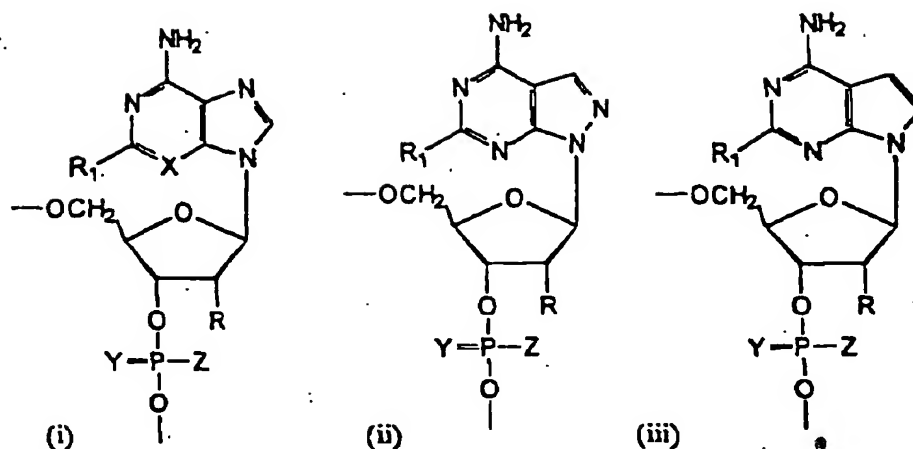
PyrroloPyr

I

Figure 4

19 MAJ 2003

Modtaget



wherein

X is N or CH;

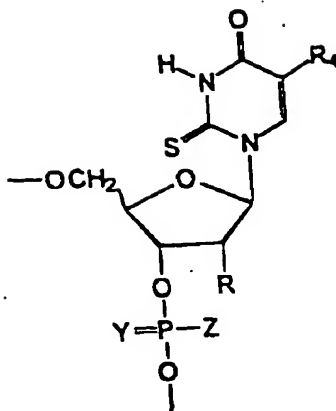
Y is O or S;

Z is OH or CH<sub>3</sub>;

R is H, F, or OR<sub>2</sub>, where R<sub>2</sub> is H, C<sub>1-4</sub> alkyl or allyl, and

R<sub>1</sub> is C<sub>1-4</sub> alkyl, C<sub>1-4</sub> alkoxy, C<sub>1-4</sub> alkylthio, F, or NHR<sub>3</sub>, where R<sub>3</sub> is H, or C<sub>1-4</sub> alkyl, and where the 8 position of the purine, the 3 position of the pyrazolopyrimidine or the 5 position of the pyrrolopyrimidine optionally serve as a point of attachment for a cross-linking function or a reporter group.

Figure 5



Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

(iv)

wherein

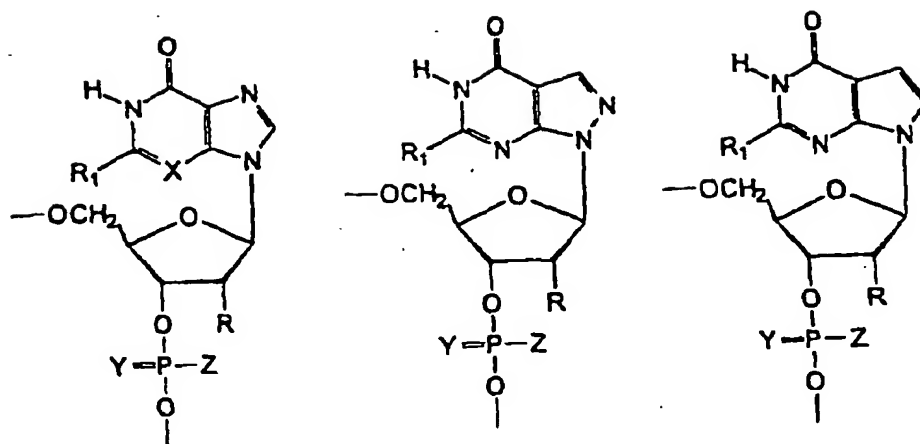
Y is O or S;

Z is OH or CH<sub>3</sub>;

R is H, F, or OR<sub>1</sub>, where R<sub>1</sub> is H, C<sub>1-6</sub> alkyl or allyl, and

R<sub>4</sub> is H, C<sub>1-6</sub> alkyl, C<sub>1-6</sub> alkenyl, C<sub>1-6</sub> alkynyl, or optionally the 5-position of the pyrimidine serves as a point of attachment for a cross-linking function or a reporter group.

Figure 6



(v)

(vi)

(vii)

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

wherein

$X$  is N or  $CH$ ;

$Y$  is O or S;

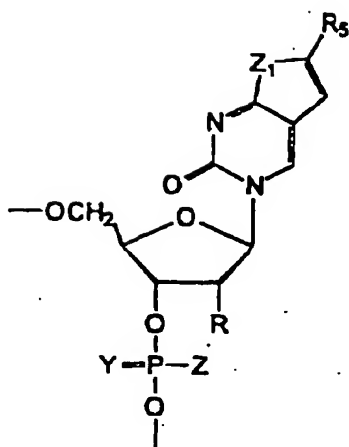
$Z$  is OH or  $CH_3$ ;

$R$  is H, F, or  $OR_2$ , where  $R_2$  is H,  $C_{1-6}$  alkyl or allyl, and

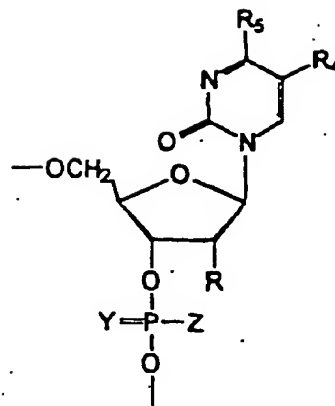
$R_1$  is H,  $C_{1-4}$  alkyl,  $C_{1-4}$  alkoxy,  $C_{1-4}$  alkylthio, F, or  $NHR_3$ , where

$R_3$  is H, or  $C_{1-4}$  alkyl, and where the 8 position of the purine, the 3 position of the pyrazolopyrimidine or the 5 position of the pyrrolopyrimidine optionally serve as a point of attachment for a cross-linking function or a reporter group.

Figure 7



(viii)



(ix)

wherein

Y is O or S;

Z is OH or CH<sub>3</sub>;

R is H, F, or OR<sub>1</sub>, where R<sub>1</sub> is H, C<sub>1-6</sub> alkyl or allyl,

R<sub>4</sub> is H, C<sub>1-6</sub> alkyl, C<sub>1-6</sub> alkenyl, C<sub>1-6</sub> alkynyl, or optionally the 5-position of the pyrimidine serves as a point of attachment for a cross-linking function or a reporter group;

Z<sub>1</sub> is O or NH, and

R<sub>5</sub> is H, or C<sub>1-6</sub> alkyl.

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

Figure 8



Patent- og  
Varemærkestyrelsen  
19 MAJ 2003  
Modtaget

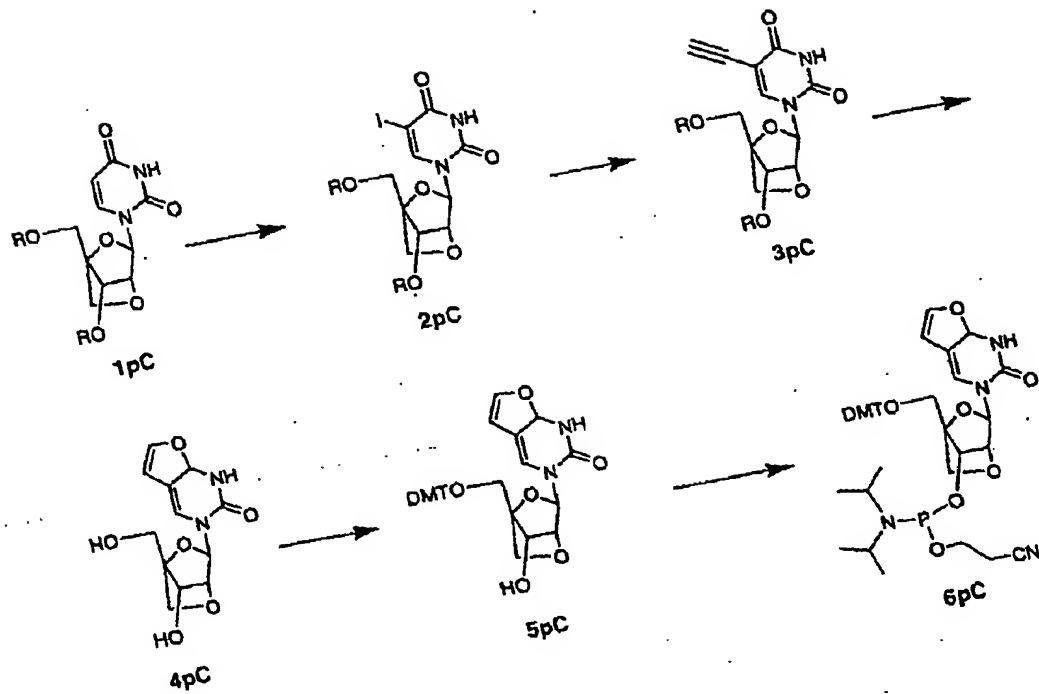


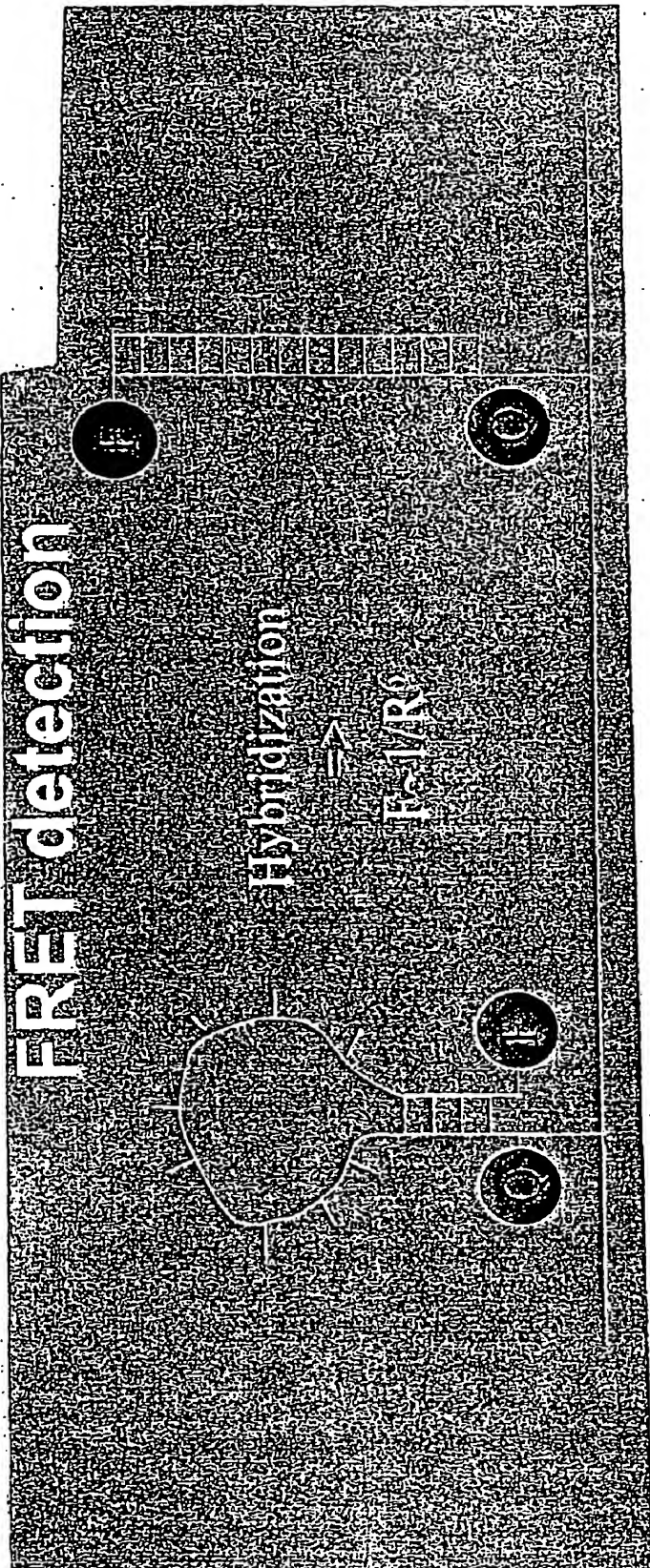
Figure 9

FRET detection

Hybridization

$\Rightarrow$

F<sub>1</sub>/R<sub>1</sub>



Patent- og  
Varemærkestyrelsen

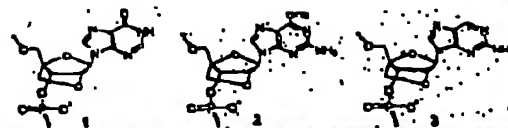
19 MAJ 2003

Modtaget

Figure 10

# BASE PAIRING PROPERTIES OF LNA (LOCKED NUCLEIC ACID) NUCLEOSIDES CONTAINING HYPOXANTHINE, 2,6-DIAMINOPURINE AND 2-AMINOPURINE NUCLEOBASES.

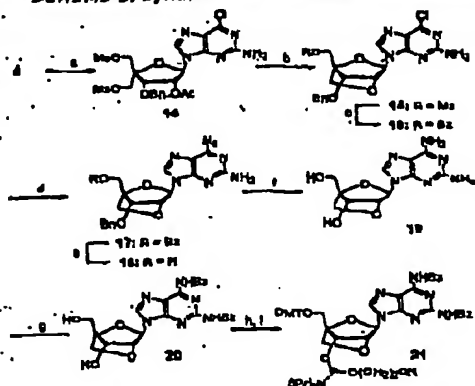
**ABSTRACT.** Three new LNA nucleosides containing hypoxanthine (LNA-I; 1), 2,6-diaminopurine (LNA-D; 2) and 2-aminopurine (LNA-ZAP; 3) nucleobases were prepared via convergent syntheses and incorporated into LNA octamers. The complexing properties of these oligonucleotides were assessed against perfect and singly mismatched DNA.



The main **OBJECTIVES** for development of second generation LNA nucleosides.

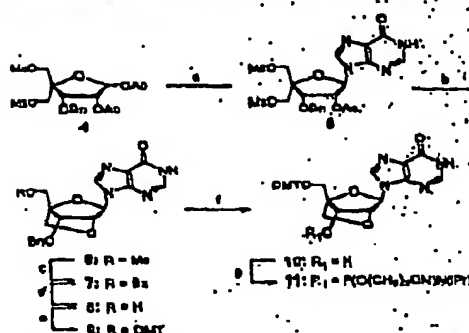
- Increased duplex stability
- Increased mismatch discrimination or universal base properties
- Equalizing duplex stability
- Recognition or selectively binding complementary (SBC) oligonucleotides

**SCHEME 2. Synthesis of the LNA-D amidite.**



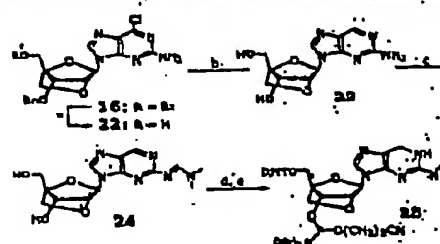
**Keys:** (a) 8-amino-2-chloropurine, BSA, TMSOTf, 1,2-dichloromethane; 90%; (b) NaOH, 1,4-dioxane, H<sub>2</sub>O; 87%; (c) NaOH, DMF; (d) NaN, DMSO; (e) NaOH, EtOH; 77% (three steps); (f) 10% Pd/C, HCO<sub>2</sub>NH<sub>4</sub>, MeOH, H<sub>2</sub>O; 84%; (g) 1. Et<sub>3</sub>N, pyridine; 2. NaOH, EtOH, pyridine; 82%; (h) DMT-Cl, pyridine; 80%; (i) 2-cyanoethyl-N,N-diisopropylphosphoramidochloridate, DIPEA, DMF; 74%.

**SCHEME 3. Synthesis of the LNA-I amidite.**



**Keys:** (a) hypoxanthine, BSA, TMSOTf, 1,2-dichloromethane; 83% (two steps); (b) NaOH, THF, EtOH, H<sub>2</sub>O; 85%; (c) NaOH, DMF; 76%; (d) NaOH, THF, MeOH, H<sub>2</sub>O; 85%; (e) DMT-Cl, pyridine; 83%; (f) Pd/C, HCO<sub>2</sub>NH<sub>4</sub>, MeOH, H<sub>2</sub>O; 77%; (g) 2-cyanoethyl-N,N-diisopropylphosphoramidochloridate, DIPEA, DMF; 75%.

**SCHEME 3. Synthesis of the LNA-ZAP amidite.**



**Keys:** (a) NaOH, 1,4-dioxane, H<sub>2</sub>O; 77%; (b) 20% NaOH/Na<sub>2</sub>CO<sub>3</sub>, MeOH, H<sub>2</sub>O; 89%; (c) N,N-dimethylformamide dimethyl acetal, DMT-Cl, pyridine; 87% (two steps); (d) 2-cyanoethyl-N,N-diisopropylphosphoramidochloridate, DIPEA, DMF; 64%.

The values of LNA/DNA and DNA/DNA duplexes. LNA nucleosides are in CAPITAL. Modified nucleosides 1 and 2 = 8-oxa-2'-phosphate, 3 and 4 = 2,6-diaminopurine nucleosides. 5 and 6 = 2-aminopurine nucleosides.

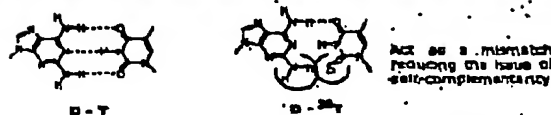
Seq.	Oligonucleotide structure	Tm (°C) of the duplex with complementary DNA	Tm (°C) of the duplex with complementary LNA	Tm (°C) of the duplex with complementary DNA	Tm (°C) of the duplex with complementary LNA
1	5'-GACGAGG	23.8	41.5	41.8	41.8
2	5'-GACGAGG	23.8	41.5	41.8	41.8
3	5'-GACGAGG	23.8	41.5	41.8	41.8
4	5'-GACGAGG	23.8	41.5	41.8	41.8
5	5'-GACGAGG	23.8	41.5	41.8	41.8
6	5'-GACGAGG	23.8	41.5	41.8	41.8
7	5'-GACGAGG	23.8	41.5	41.8	41.8
8	5'-GACGAGG	23.8	41.5	41.8	41.8
9	5'-GACGAGG	23.8	41.5	41.8	41.8
10	5'-GACGAGG	23.8	41.5	41.8	41.8
11	5'-GACGAGG	23.8	41.5	41.8	41.8
12	5'-GACGAGG	23.8	41.5	41.8	41.8
13	5'-GACGAGG	23.8	41.5	41.8	41.8
14	5'-GACGAGG	23.8	41.5	41.8	41.8

The melting temperatures (Tm values) were determined as the midpoint of the sigmoidal melting curves (midpoint at 50% duplex formation). Concentration of the duplexes: 2.5 μM. Buffer: 0.1 M NaCl; 10 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 7.2); 1 mM EDTA.

Table 1 (continued)

1. Kuchta, R. P.; Kuchta, R. P.; Kuchta, R. P.; Kuchta, R. P. *J. Org. Chem.* 1993, 58, 6204.  
2. Kuchta, R. P.; Kuchta, R. P.; Kuchta, R. P.; Kuchta, R. P. *J. Org. Chem.* 1993, 58, 6204.  
3. Kuchta, R. P.; Kuchta, R. P.; Kuchta, R. P.; Kuchta, R. P. *J. Org. Chem.* 1993, 58, 6204.

The principle of SBC oligonucleotides construction [3].



oligonucleotide structure	Tm (°C) of complementary duplexes with (where a = 2-aminopurine)				
	3'-GACGAGG	3'-GACGAGG	3'-GACGAGG	3'-GACGAGG	3'-GACGAGG
5'-GACGAGG	23.8	41.5	41.8	41.8	41.8
5'-GACGAGG	23.8	41.5	41.8	41.8	41.8
5'-GACGAGG	23.8	41.5	41.8	41.8	41.8
5'-GACGAGG	23.8	41.5	41.8	41.8	41.8

\* Tm values in the second column were determined in the same buffer as in the first column, but with the cooperativity of the second strand (sequence 41.8°C).

Patent: og  
Varemærkestyrelsen

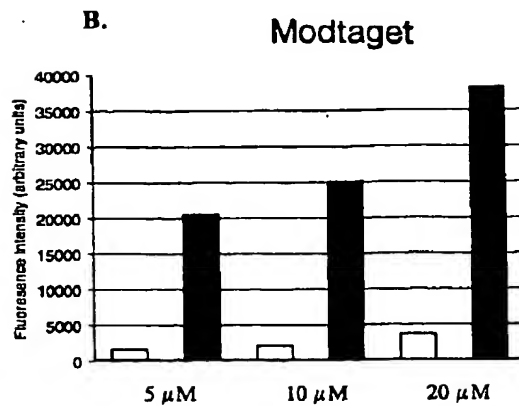
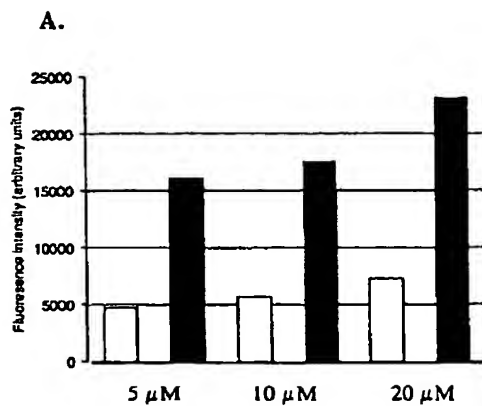
19 MAJ 2003

Modtaget

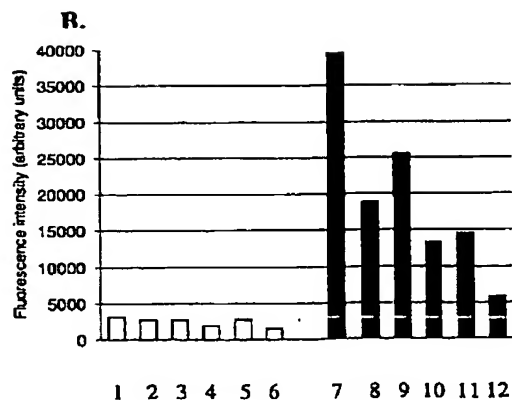
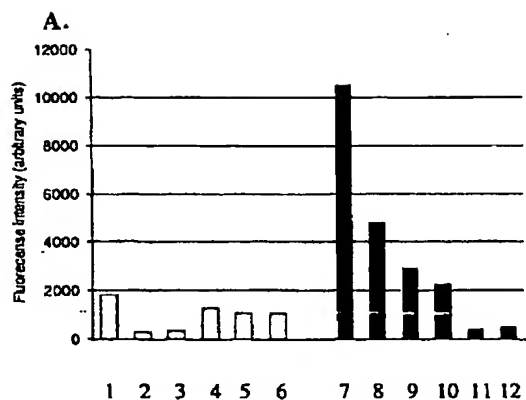
Figure 11

19 MAJ 2003

Modtaget

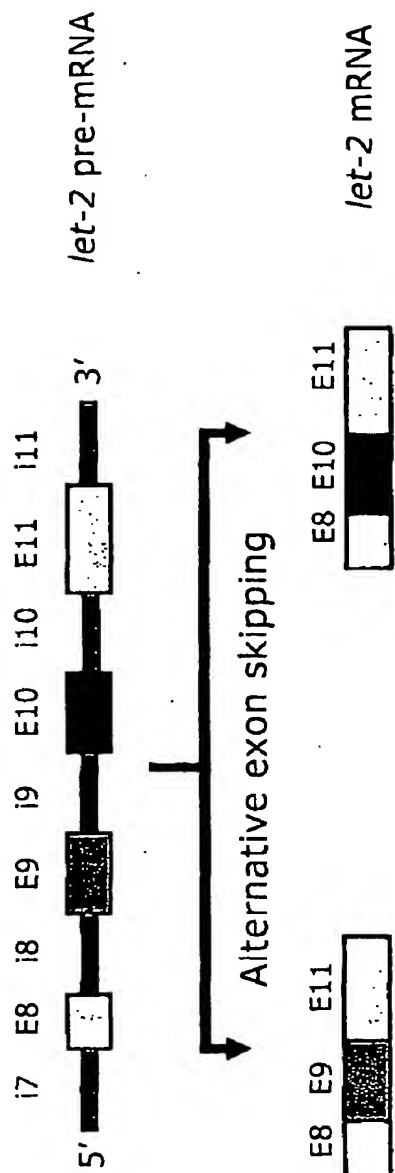


*Figures 12A and 12B*



*Figures 13A and 13B*

# Detection principle for *C. elegans* Let-2



Cy3-labeled cDNA from embryos      Cy5-labeled cDNA from L1 larvae

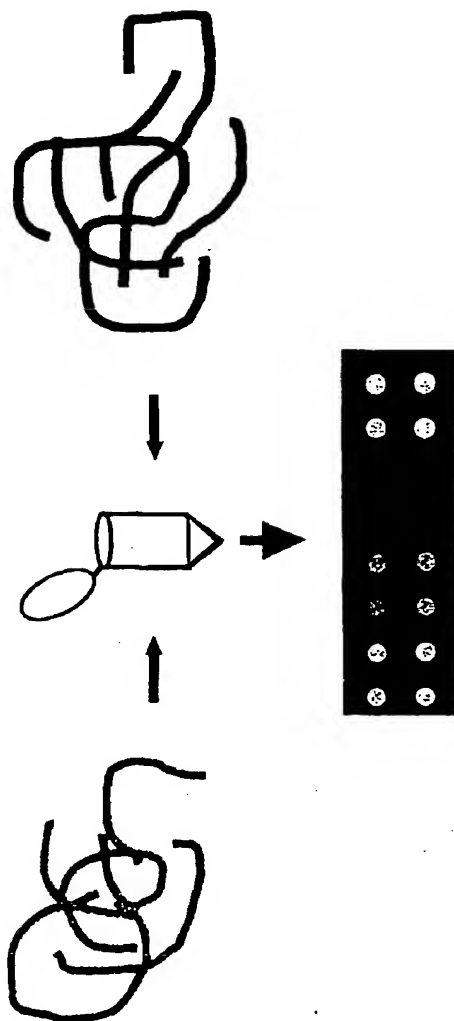
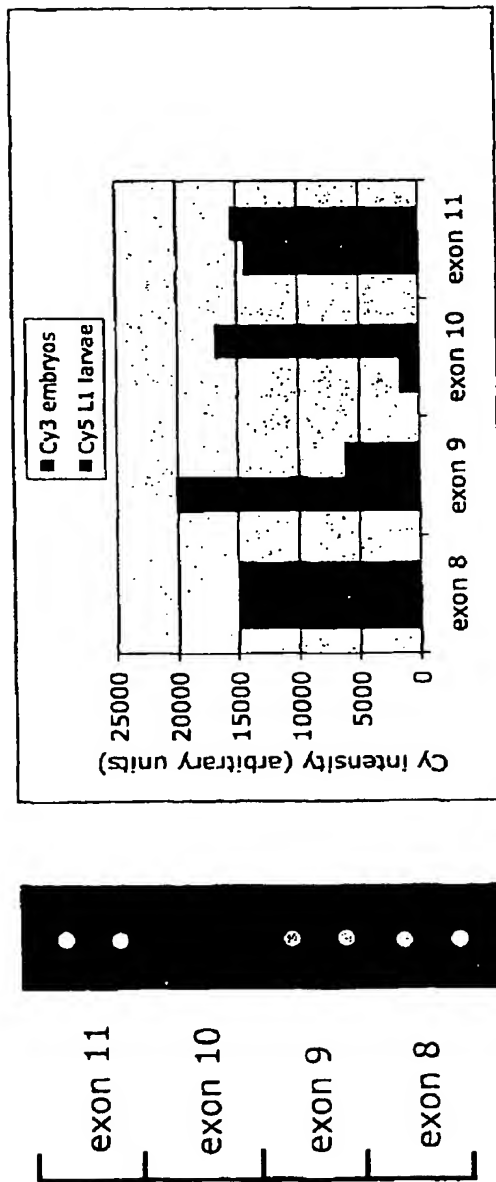


Figure 14

# Alternative splicing in the *C. elegans* *Let-2* mRNA

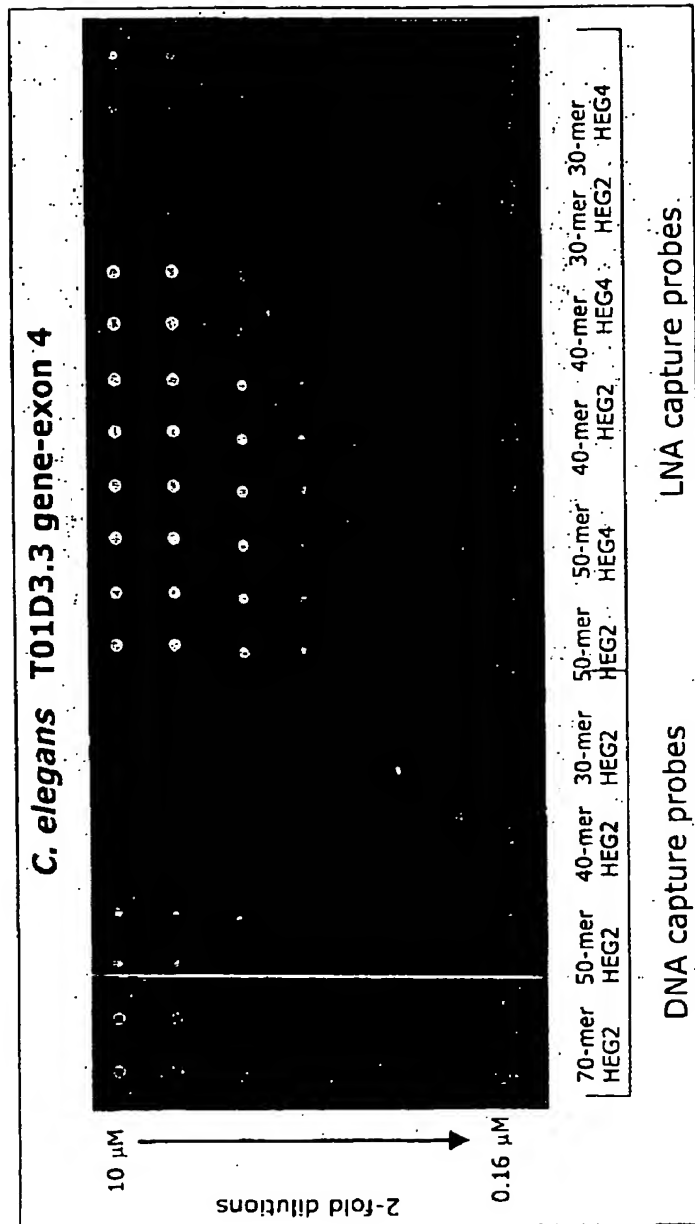


Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

Figure 15



Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

Figure 16A

19 MAJ 2003

Modtaget

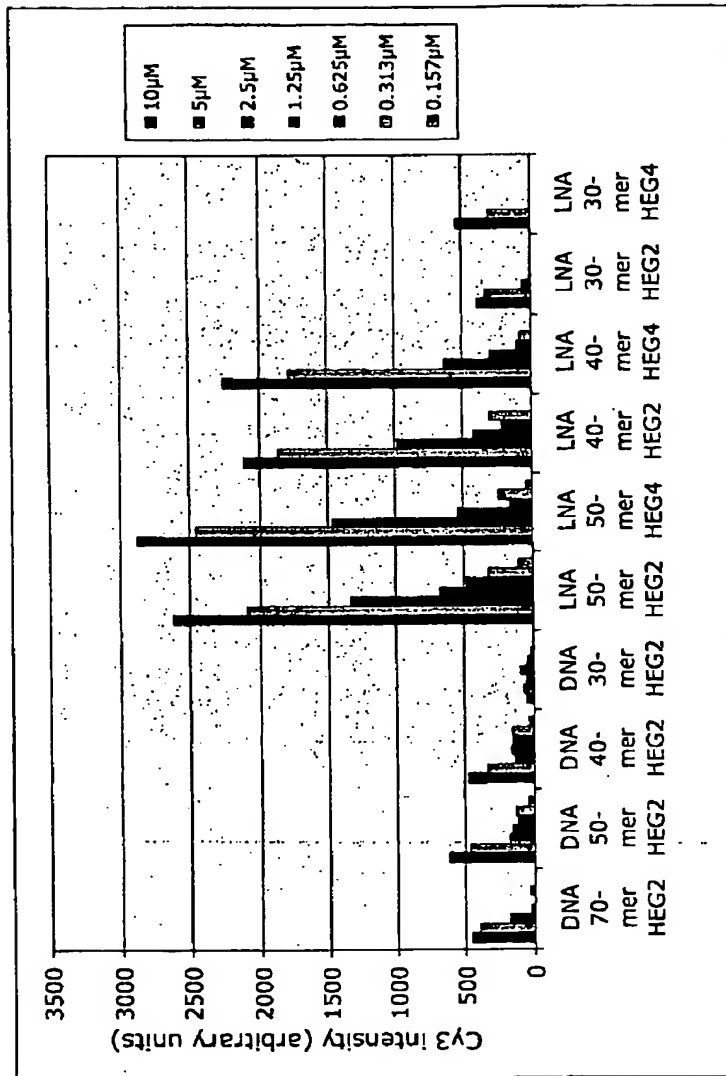


Figure 16B



Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

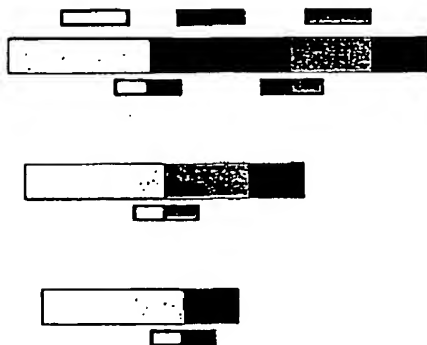


Figure 17

19 MAJ 2003

Modtaget

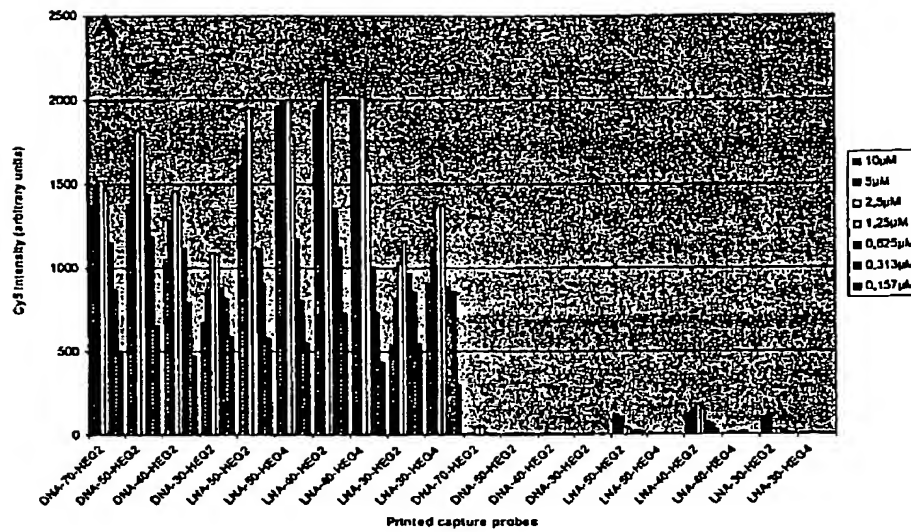


Figure 18A

19 MAJ 2003

Modtaget

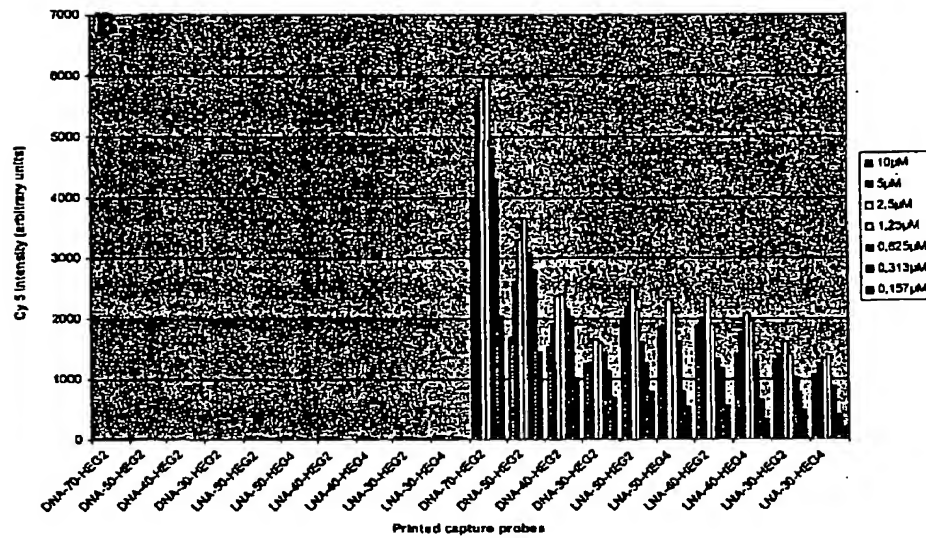


Figure 18B

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

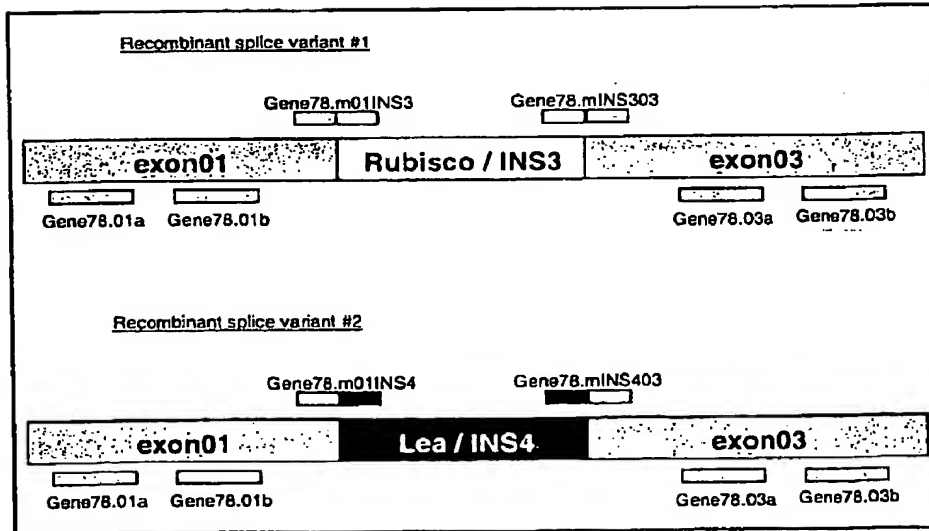


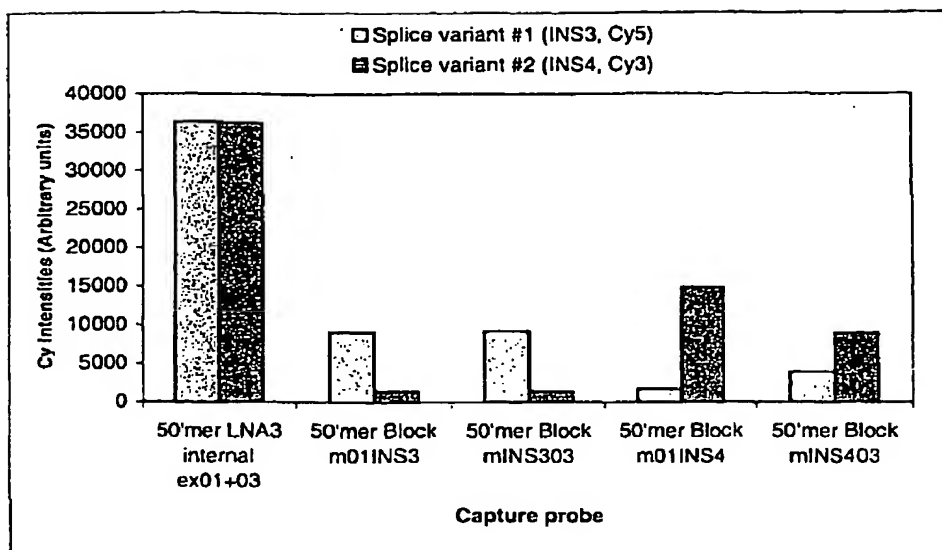
Figure 19

A

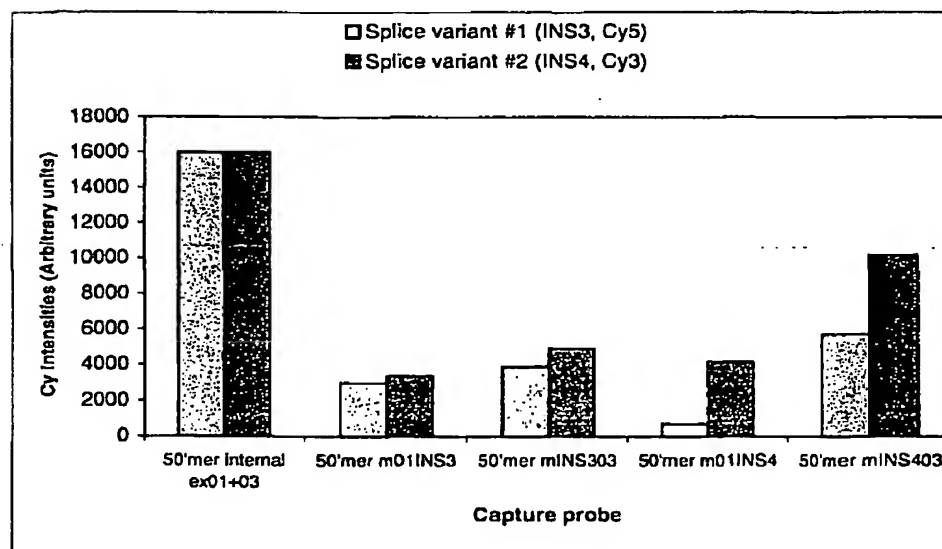
Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget



B



Figures 20A and 20B

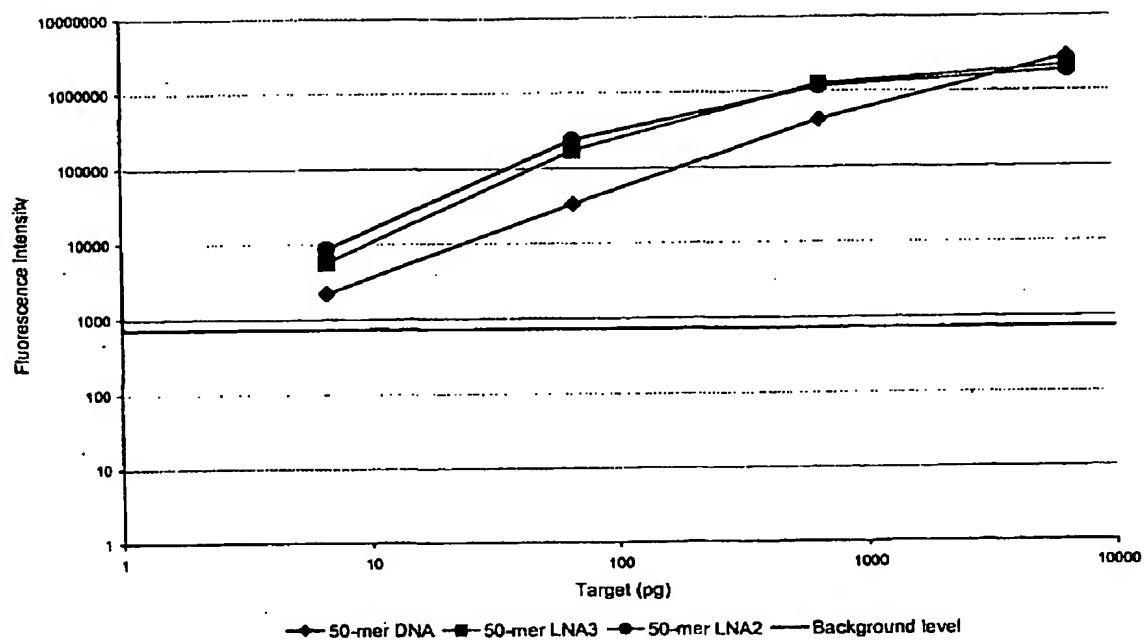


Figure 21

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

19 MAJ 2003

Modtaget

Signal  
Background

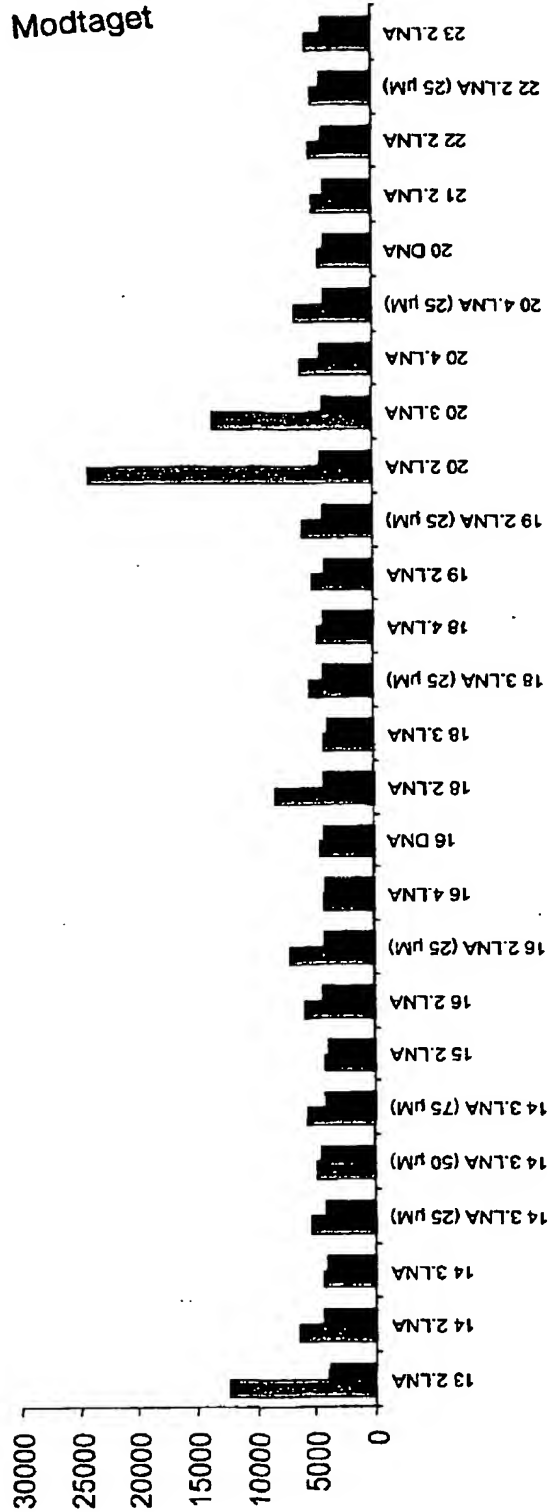
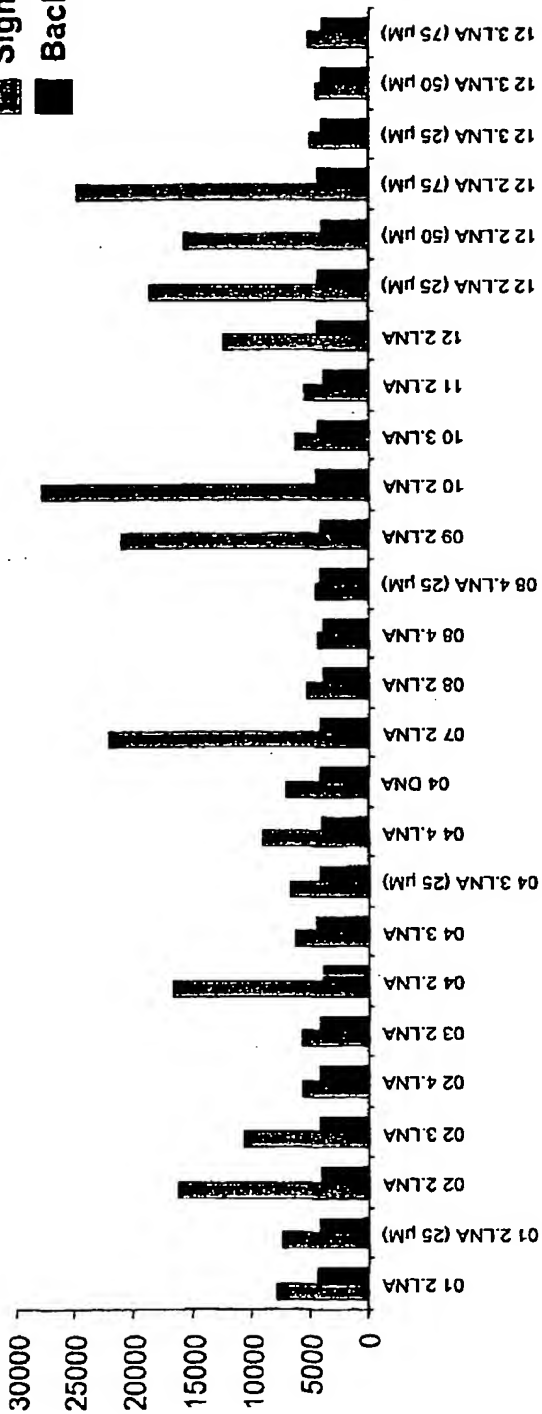
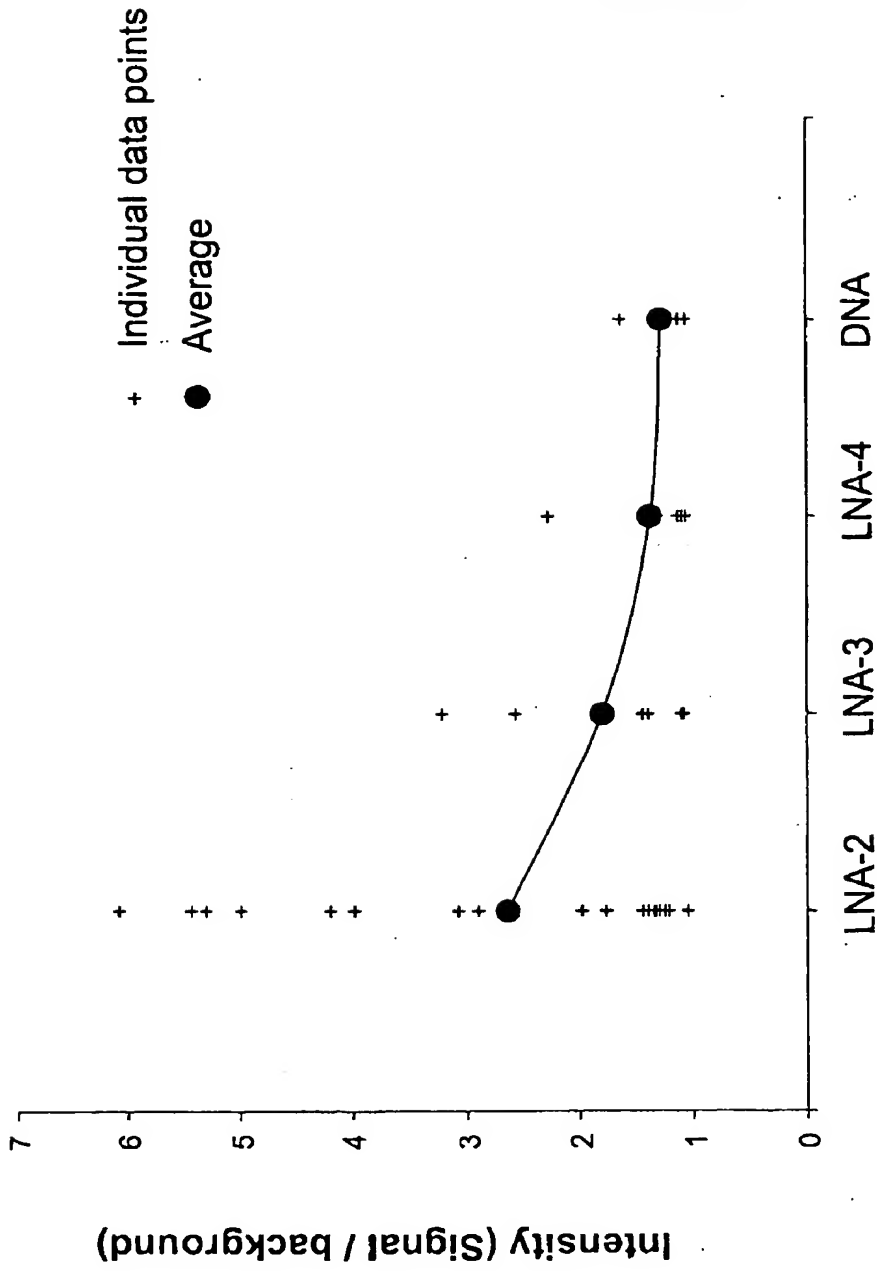


Figure 7

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget



Frequency of LNA monomers in oligonucleotides

Figure 23



19 MAJ 2003

Modtaget

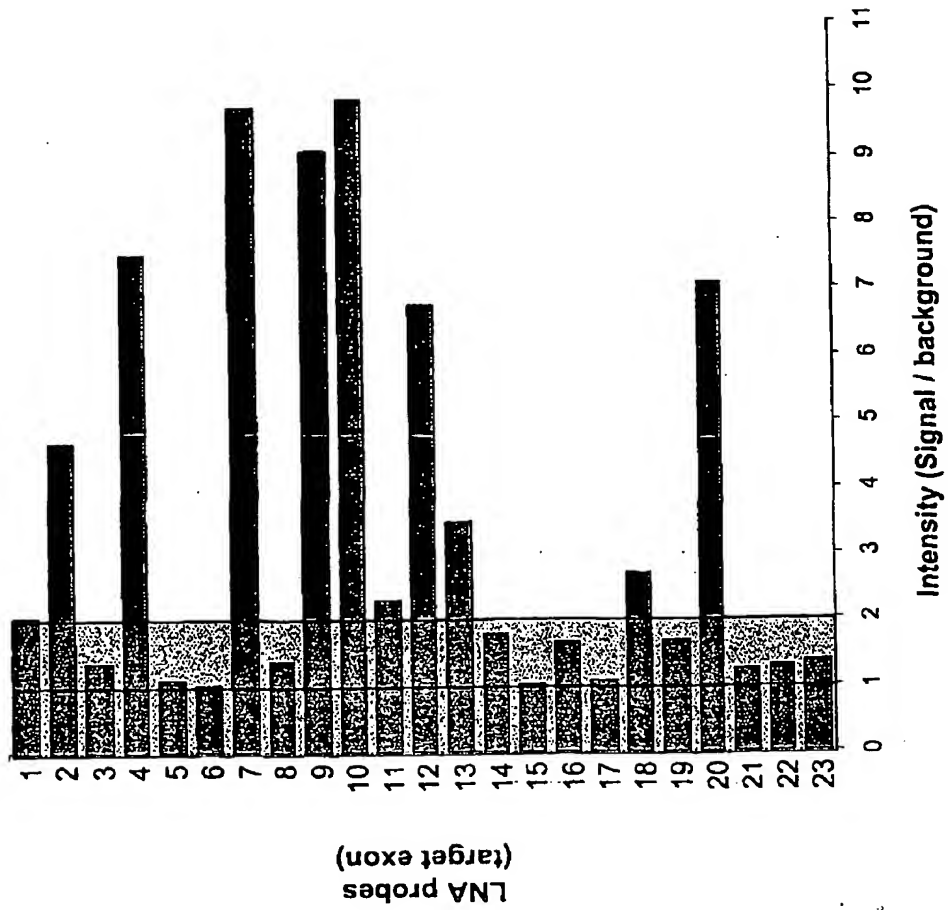


Figure 24

Figure 25

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

EO No	Oligo Name	Sequence
8253	Menkes.02 50NH2C6-DNA	tcgttgagggtatgacttgcaattccctgtgttggaaccattgagcagca
8254	Menkes.02 50NH2C6-4.LNA	TctgTtgaGggtAtgamCttgmCaatTcctGtgtTtggAccaTtgaGcagmCa
8255	Menkes.02 50NH2C6-3.LNA	TctGtGagGgtAtgActTgcAatTccTgtGtTggAccAttGagmCagmCa
8256	Menkes.02 50NH2C6-2.LNA	TcTgTtGaGgGtAtGamCttgmCaAtTcmCtGtGtTtGgAcnCaTtGaGcAgmCa
8258	Menkes.04 50NH2C6-DNA	agaaaagcaatagaggctgtatcacccggggctatatagagttatcat
8259	Menkes.04 50NH2C6-4.LNA	AgaaAagcAataGaggmCtgtAtcamCccgGgctAtatAgagTtagTatcAc
8260	Menkes.04 50NH2C6-3.LNA	AgaAaaGcaAtaGagGctGtaTcamCcgGggmCtaTatAgaGttAgtAtcAc
8261	Menkes.04 50NH2C6-2.LNA	AgAaAaGcAaTaGaGgmCtGtAtmCamCcGgGgmCtAtAtAgAgTtAgTaTcAc
8263	Menkes.06 50NH2C6-DNA	gctgttatacaacccccaatgatagcagagttcatccgagaacttgatt
8264	Menkes.06 50NH2C6-4.LNA	GctgTtatAcaamCcccmCaatGataGcagAgttmCattcmCgagAactTggaTt
8265	Menkes.06 50NH2C6-3.LNA	GctGtAtamCaamCccmCcaAtgAtaGcaGagTtcAtcmCgaGaamCttGgaTt
8266	Menkes.06 50NH2C6-2.LNA	GcTgTtAtAcAamCcmCcmCaAtGaTaGcAgAgTtmCaTcmCgAgAamCttGgAtT
8268	Menkes.08 50NH2C6-DNA	tcttgggtcaagaaggatcgggtcagcaagtcacttagatcataaacgaga
8269	Menkes.08 50NH2C6-4.LNA	TcttTggtmCaagAaggAtcgGtcaGcaaGtcmCttaGatcAtaaAcgaGa
8270	Menkes.08 50NH2C6-3.LNA	TctTtGtGcAagAagGatmCggTcaGcaAgtmCacTtaGatmCatAaamCgaGa
8271	Menkes.08 50NH2C6-2.LNA	TcTtTgGtmCaAgAaGgAtmCgGtmCaGcAaGtmCamCtTaGaTcAtAaAcGaGa
8273	Menkes.10 50NH2C6-DNA	tataaagcactgaagcataagacagcaaatatggacgtactgattgtgc
8274	Menkes.10 50NH2C6-4.LNA	TtatAaagmCactGaagmCataAgacAgcaAataTggamCgtamCtgaTgtGc
8275	Menkes.10 50NH2C6-3.LNA	TtaTaaAgcActGaaGcaTaaGacAgcAaaTatGgamCgtActGatTgtGc
8276	Menkes.10 50NH2C6-2.LNA	TtAtAaAgmCamCtGaAgmCaTaAgAcAgmCaAaTaTgGamCgtamCtGaTtGtGc
8278	Menkes.12 50NH2C6-DNA	aacaagtggaigtggaactgtacaacgtggagatatcatlaaagtagtt
8279	Menkes.12 50NH2C6-4.LNA	AacaAgigGalTggaActtGtacAacgTggaGataTcatTaaaGtagTt
8280	Menkes.12 50NH2C6-3.LNA	AacAagTggAtgTggAacTtgTacAacGtgGagAtaTcaTtaAagTagTt
8281	Menkes.12 50NH2C6-2.LNA	AamCaAgTgGaTgTgGaAcTtGtAcAamCgTgGaGaTaTcatTaaGtAgTt
8283	Menkes.14 50NH2C6-DNA	ccattgccaccctctgtgtatggattgtaattggatttctgaattttgaa
8284	Menkes.14 50NH2C6-4.LNA	mCcatTgccAccctCttGgtaTggaTtgtAattGgatTtctGaatTtga
8285	Menkes.14 50NH2C6-3.LNA	mCcaTtgmCcamCccTctTggTatGgaTtgTaaTtgGatTtcTgaAttTgAa
8286	Menkes.14 50NH2C6-2.LNA	mCcAtTgmCcAcmCcTcTtGgTaTgGaTtGtAaTtGgAtTtmCtGaAtTtGtAa
8288	Menkes.16 50NH2C6-DNA	ggtatttgalaagactggaaccattactcacgggaaccccaatggtgaatc
8289	Menkes.16 50NH2C6-4.LNA	GgtaTtgaAaaGactGgaamCcatTactmCacgGaaamCccaGtggTgaaTc
8290	Menkes.16 50NH2C6-3.LNA	GgtAttTgaTaaGacTggAacmCatTactcamCggAacmCccAgtGgtGaaTc
8291	Menkes.16 50NH2C6-2.LNA	GgTaTtTgAaAaGamCtGgAamCcAtTamCtmCamCgGaAcmCcmCaGtGgTgAaTc

# Figure 25 (continued)

8293 Menkes.18 50NH2C6-DNA  
8294 Menkes.18 50NH2C6-4.LNA  
8295 Menkes.18 50NH2C6-3.LNA  
8296 Menkes.18 50NH2C6-2.LNA

8298 Menkes.20 50NH2C6-DNA  
8299 Menkes.20 50NH2C6-4.LNA  
8300 Menkes.20 50NH2C6-3.LNA  
8301 Menkes.20 50NH2C6-2.LNA

altggaacogggagtggaatgataaaggcttgcattaataacga  
AttgGtaamCoggGagtGgatGattAgaaAiggTcttGtcaTtaaTaacGa  
AltGgtAacmCoggGagTggAigAltAgaAatGgtmCttGtcAttAatAacGa  
AltGgGtAamCcGgGaGtGgAltGatTAgaAaAtGgTcTtGtmCaTIAaTaAcGa

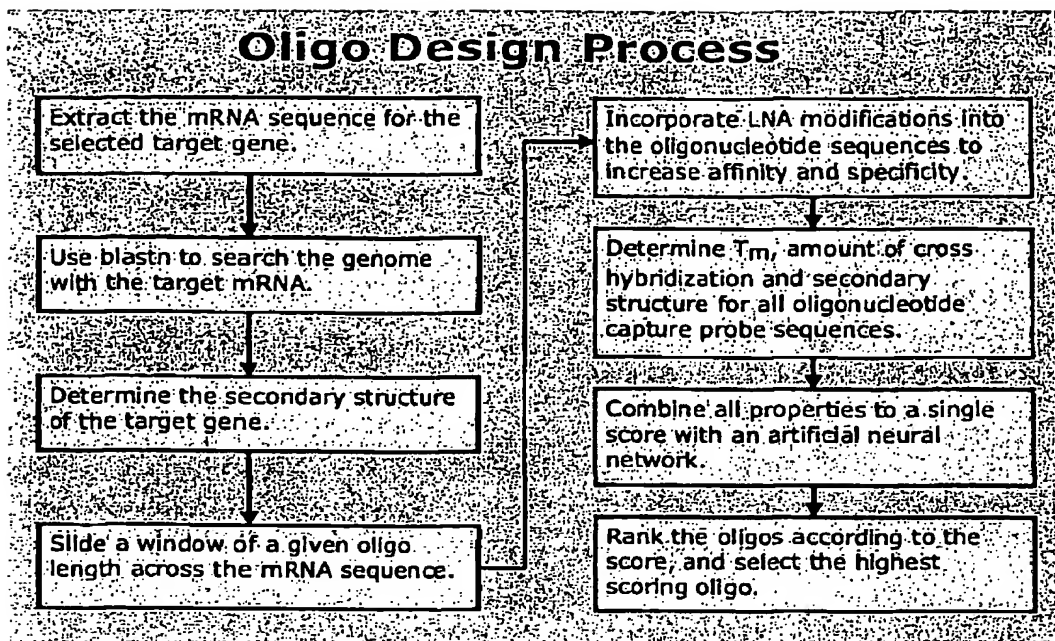
tggcacaggcacagatgtagccattgaagcagctgatgtgtttgataa  
TggcAcaggGcacAgatGtagmCcaTtgaaGcagmCtgaTgtgGttTgatAa  
TggmCacAggmCacAgaTgtAgcmCalTgaAgcAgcTgaTgtGgtTtGatAa  
TgGcAcAgGcAcAgAltGtAgmCcAltTgaAaGcAgmCtGaTgtGgtTtGatAa

## EO No. Oligo Name

10573 Menkes.01 50NH2C6-2.LNA  
10574 Menkes.03 50NH2C6-2.LNA  
10575 Menkes.05 50NH2C6-2.LNA  
10576 Menkes.07 50NH2C6-2.LNA  
10577 Menkes.09 50NH2C6-2.LNA  
10578 Menkes.11 50NH2C6-2.LNA  
10579 Menkes.13 50NH2C6-2.LNA  
10580 Menkes.14 50NH2C6-2.LNA  
10581 Menkes.15 50NH2C6-2.LNA  
10582 Menkes.17 50NH2C6-2.LNA  
10583 Menkes.19 50NH2C6-2.LNA  
10584 Menkes.21 50NH2C6-2.LNA  
10585 Menkes.23 50NH2C6-2.LNA  
10705 Menkes.22 50NH2C6-2.LNA

## Sequence

GtGmCtTcTcmCgAltTgTgAgmCtTtGtGgGcmCtGcGtAcGtGgAltT  
TtTtAamCtGmCamCcTtGtTmCtGmCtGtTmCgGcGtmCamCtGmCtTtGcmCa  
mCaTmCaGgTcAcTgGcAltGmCtTgmCgmCtTcmCtGtGtAgmCaAamCaTtGgAc  
TgAgGgGaAltGmCgTgTgmCcTcmCtGcGtAcAtAaAaTaGaGtmCtAgTcTc  
TgTaTmCcTgTaAltGgGgmCtGaTgAcAltAltGgTgGtTaTgGmCcAcmmCa  
AcAltCaGaGgmCtmCtTgmCaAaGtTaAltTmCamCtAcAaGcTmCaGaAgmCaAc  
TmCcAltAaCmCaGaAcGgGtmCamCtGcTtAltCtGcGcAamCamCaTgTtGgAg  
mCcAltGmCcAcmmCcTcTtGgTaTgGaTtGtAaTtGgAltTmCtGaAltTtGgAa  
GaAamCgAltAaTamCgAltTgCtTtTcmCaAgmCcTcTaTcAcAgTmCtGtTgmCa  
AltGaAcAgTcAltCaCtTmCgTcTmCcAltGtAltTgAltGcmCcAgAltCtmCa  
GtTcTgAltGmCtGgAgAcAamCaGtAaAamCaGcTaGaTcTaTtGcTmCtmCa  
TgGcAaGtAltGtAltTmCaAgAaAgAcAgTcAaGaGgAltCgGgAltAaAt  
GcmCtmCtAltAaAcTcAcTmCtGtmCtGaTaAamCgmCtmCcmCtAaAcAgTgTtGt  
mCtGgAltGgGaTcTgmCaGcAaTgGcTgmCtTcAltCtGtTmCtGtAgTamCtTt



Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

Figure 26

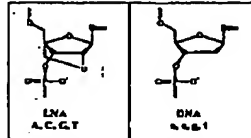
## Oligo Design Considerations

Gene expression profiling using oligo micro arrays has become a central technology in pharmacogenomics and systems biology. Sensitivity and specificity are key elements in the development of oligo probes for micro arrays. A high sensitivity can be achieved by using LNA (Locked Nucleic Acid) modified oligos spotted onto a polymer slide. A high specificity can be reached through optimal oligo design.

The *OmniDesign* system features LMA modified oligonucleotide secondary structure prediction, LMA spliced RNA melting temperature prediction, genome-wide cross hybridization prediction, secondary structure prediction of the target and recognition and flanking of the target in the genome. These features are determined for each possible probe of the query gene and presented to an artificial neural network. The probes are heuristically ranked according to the neural network prediction and the top scoring probes are returned.

The OligoDesign tools can be used freely at <http://fastools.com/>.

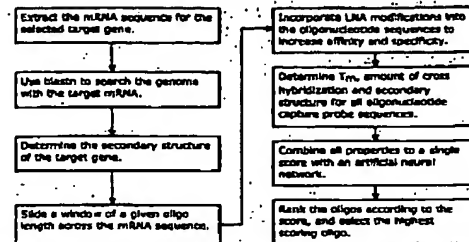
**LNA**  
- A novel  
DNA analog



Locked Nucleic Acid is a novel chemical DNA analog, that can be substituted in DNA and RNA sequences. LNA obeys the Watson-Crick pairing rules, but has increased specificity and affinity in hybridization. Designing mixed LNA/DNA oligonucleotides facilitates a control of the melting properties of probes and primers, and an optimization of sequence discrimination at a given T.

- **Uniqueness of edge**
  - No cross hybridization towards other genes.
- **Avoidance of secondary structure**
  - Self folding of the edge.
  - Target folding.
- **Splicing with modified DNA analogues using LNA**
  - Increased specificity and affinity towards the target gene.
- **Tm prediction for LNA spiked edges**
  - Control the melting properties ( $T_m$ ) within specified limits

## Oligo-Design Process



### Cross hybridization

Possible cross hybridization to other genes can be detected from a blast search against the genome.

```
TGATGGCTTCCCAATTAGGTCTTCAC
```

	18 matches
CCTCACAATTGTTGCCTCCTC	Chromosome_III
	6 matches
CGGAAC	Chromosome_I

## Secondary structure prediction

- Based on Nussinov algorithm.
- High scoring substructures detected only.
- Scoring scheme derived from melting experiments with LNA modified oligonucleotides...

TTCTATCATAAGAGAAATCTTGTGAATATC  
((( ( ) ( . ( ) ) . ) ) ) )

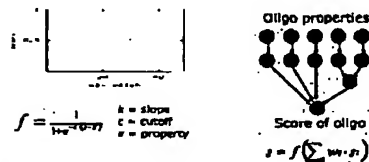
**Secondary structure of an LNA spliced oligo.**

[illegible]

Regions in the target game with high secondary structure content are masked.

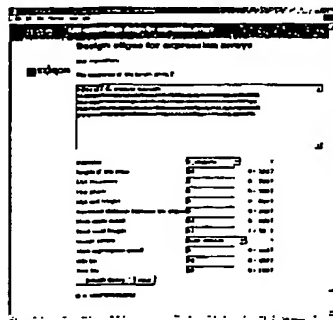
## Oligo Scoring Using Artificial Neural Network

For all possible oligos a score representing how well the oligo fits the design criteria is computed.

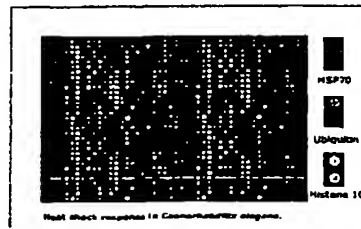


The first layer of the network applies a squashing function  $f$  with a specified slope and cutoff to every oligo property. The second layer combines the properties by a weighted sum  $s$ . Currently the network parameters have been hand tuned, as more data become available it will be possible to train the network with backpropagation to optimise the ranking of oligos.

**www.inatools.com**



### C. elegans heat shock Expression profile



- Comparative hybridization with Cy3 (standards) and Cy5 (heat treated) labelled cDNA from L4/adult *C. elegans* RNA
- Typical heat shock response for the HSP70 gene
- Expression of Ubiquitin is unaffected whereas expression of Histone H2A is repressed

## Conclusion

With the ability to increase the melting temperature and affinity towards DNA it is possible to design UNA spiked oligos with balanced Tm for Micro Array Gene Expression Profiling.

The OligoDesign system selects unique oligos with high specificity towards the target, even Tm across the array slide and decreased secondary structure. This ensures sensitivity and robustness and allows monitoring of gene expression otherwise not possible.

Figure 27

Figure 28A

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

1

```
#!/usr/bin/perl -w
#$Revision: 1.18 $
#
# $Id: oligod,v 1.18 2002/07/29 06:26:35 nt Exp $
5 #
# NAME
# oligod - microarray oligod design
#
# SYNOPSIS
10 #   oligod [-blastdb blastdb] [-fastadb fastadb] fastaseq
#       [-length oligolength]
#       [-mindist minimunoligodistance]
#       [-maxhits n]
#       [-min_tm t]
15 #       [-max_tm t]
#       [-maxoligo n]
#       [-minscore n]
#       [-gene_ident_cutoff ident]
#       [-gene_len_cutoff len]
20 #       [-comp|-revcomp|-rev]
#       [-capitalize -cphase n -cfreq n]
#       [-verbose]
#       [-self_score]
#       [-param] parameterfile
25 #       [-conf] configurationfile
#       [-lna]
#       [-matrix] hybridisationmatrix
#
# DESCRIPTION
30 #   1. Blast against blastdb
#   2. Less than 17 consecutive matching nt's
#   3. Identity less than 60%
#   4. Palindrom = SmithWaterman against complement
#   5. Melting temperature  $(0.88 \times A/T) + (1.47 \times G/C) + (X)$ 
35 #       [PICK70 script (DeRisi group)] (+X is by Niels
Tolstrup)
#   6. Salt:
#       (http://jsll1.chem.wayne.edu/Hyther/hytherm2main.html)
#       a. Monovalent cation 1 mol/L
40 #       b. Mg2+ 0 mol/L
#       c. Hybridization temperature 37.0 degrees Celcius
#       d. Target 2e-7 mol/L
#       e. Primer 1e-9 mol/L
#   7. Number of primers to show.
45 #   8. Primer length.
#
#       -maxhits n
#       n is the maximal number of alignments to show for each oligo
#
50 #       -maxoligo n
#       n is the maximal number of oligos to suggest for each
sequence
#
#       -min_score n
55 #       n is the minimal score required for a hit to be included in
the
#       scoring of a oligo
#
```

```

# -mindist n
# Oligos will not overlap if n = oligolength
# 0 and 1 allow for all oligos
#
5 # -gene_ident_cutoff ident
# Mask hits with identity larger than or equal to ident.
# Used together with gene_len_cutoff.
# Default 98
#
10 # -gene_len_cutoff len
# Mask hits with alignment length longer than or equal to len
# Used together with gene_ident_cutoff.
# Default 50
#
15 # -lna
# If a LNA matrix is used for selfhybridisation this must be
# set.
#
# -matrix hybridisationmatrix
20 # The format of the hybridisation matrix is that used by the
# fasta
# package. LNA is represented by A=L C=I G=O and T=U in the
# matrix.
#
25 # -capitalize
# -cphase phase
# -cfreq freq
# -end_spike_len len
# phase is 0 to freq - 1
30 # freq is the frequency of LNA default is 4
#
# -blastdb "db1 db2 db3"
#
# -sense reverse
35 # Uses the reverse complement of the query, use with strand =
# bottom
#
# oligod reads the configuration file oligod.conf
#
40 #FILES
# blastall
# formatdb
# dan
# lynx
45 # ssearch
# dyp
#
#AUTHOR
# Copyright Niels Tolstrup 2001,2002
50 # tolstrup@exiqon.com
# Exiqon
#
use strict;
55 use Getopt::Long;
use FindBin qw($Bin);
use lib "$Bin/./lib";

```

```

use LWP::UserAgent;
use HTTP::Cookies;

5 my $glb;

$glb->( dir)          = $Bin;
$glb->( conf)         = "$glb->( dir)/oligod.conf";
10 $glb->( tmpdir)      = "$glb->( dir)/tmp";
$glb->( program_name) = "$glb->( dir)/oligod";
$glb->( lynx_cookie)   = "$glb->( dir)/lynx_cookies";
$glb->( dyp)           = "$glb->( dir)/bin/dyp";
$glb->( bin)           = "/usr/bin";
15 $glb->( formatdb)    = "$glb->( bin)/formatdb";
$glb->( blastall)      = "$glb->( bin)/blastall";
$glb->( ssearch)       = "$glb->( bin)/ssearch";
$glb->( dan)           = "$glb->( bin)/dan";
$glb->( lynx_env)      = "LD_LIBRARY_PATH=/usr/local/lib ";
20 $glb->( lynx)        = "$glb->( lynx_env)$glb->( bin)/lynx";
$glb->( tm_exiqon_url) = "http://armstrong/cgi-
bin/tmpredict.cgi";
$glb->( tm_url)        = 'http://lna-tm.com/';
$glb->( ua)            = LWP::UserAgent->new;
25 $glb->( cookiefn)    = "$glb->( dir)/lwpcookies.txt";
$glb->( ua)->cookie_jar( HTTP::Cookies->new( file => $glb->(
cookiefn),
                                autosave => 1));

30

$glb->( version)      = "0.9";
$glb->( verbose)      = 0;
#$glb->( view)        =
35 "BLAST:FASTA:SELF_EVIDENCE:PROBE:RUN_TIME:PARAMETER" .
   ":SCORE_EVIDENCE";
#$glb->( view)        =
"BLAST:SELF_EVIDENCE:PROBE:RUN_TIME:SCORE_EVIDENCE";
$glb->( view)         =
40 "PARAMETER:FASTA:BLAST:SELF_EVIDENCE:PROBE:RUN_TIME" .
   ".:SCORE_EVIDENCE:RUNTIME:TARGET_STRUCT";
#$glb->( view)        = "";

45 ##### XML I/O #####
#use XML::Dumper;
#sub dump_parameter_xml($$){
# my ( $fn, $dat) = @_;
# my $dump = new XML::Dumper;
50 # my $xml = $dump->pl2xml( $dat);
# local FILE;
# open( FILE, "> $fn") || die "Could not open $fn !";
# print FILE "$xml\n";
# close FILE;
55 #}
#
#sub read_parameter_xml($){
# my ( $fn) = @_;

```



```

# my $xml = "";
# local FILE;
# open( FILE, "< $fn" ) || die "Could not open $fn $!";
# while( <FILE> ){
5 #   $xml .= $_;
# }
# my $dump = new XML::Dumper;
# my $dat = $dump->xml2pl( $xml );
# close FILE;
10 # $dat;
#)

##### PARAMETER I/O #####

15 sub read_conf($){
    my ( $glb ) = @_;
    my $var;
    local *FILE;

20   open( FILE, "$glb->( conf )" ) || die "Could not open $glb->(conf)";
    while ( <FILE> ) {
        chomp;
        if( /^s*#/ ) { }
        elsif( /^s*$/ ) { }
25        elsif( /^s*(\S+)\s+([^\s]+)\s*/ ) {
            if( defined $glb->( $1 ) ) {
                $glb->( $1 ) = $2;
            }
30        }
        else {
            print "$_ was not recognized\n";
        }
    }
35   close FILE;
    foreach $var ( "tmpdir", "lynx_cookie" ) {
        if( $glb->( $var ) =~ /\s+/ ) {
            $glb->( $var ) = "$glb->( dir )/$glb->( $var )";
40        }
    }
}

sub read_param($$){
45   my ( $param, $file ) = @_;
    my $warning = '';
    my $tmp1;
    my $tmp2;

50   while ( <$file> ) {
        if( /^s*#/ ) { }
        elsif( /^s*$/ ) { }
        elsif( /^s*<PARAMETER>$/ ) { }
        elsif( /^s*</PARAMETER>$/ ) {
55         last;
        }
        elsif( /^s*oligod_param/i ){
            if( /^s*oligod_param\s+(\S+)\s+(\S+)\s+(\S+)/i ) {

```

```

$tmp1 = $1;
$tmp2 = $2;
$tmp1 =~ tr/A-Z/a-z/;
$tmp2 =~ tr/A-Z/a-z/;
5  if( defined $param->(oligod_param){ $tmp1}) {
    if( defined $param->(oligod_param){ $tmp1}{ $tmp2}) {
      $param->(oligod_param){ $tmp1}{ $tmp2} = $3;
    }
    else {
10    $warning .= "Did not recognize $tmp2 in $_";
    }
    else {
15    $warning .= "Did not recognize $tmp1 in $_";
    }
    else {
      $warning .= "Not enough values fore oligod_param in $_";
    }
20 }
elseif( /^s*blast_param/i){
  if( /^s*blast_param\s+(\S+)\s+(.+)/i ) {
    $tmp1 = $1;
    $tmp1 =~ tr/A-Z/a-z/;
25    if( defined $param->( blast_param){ $tmp1}) {
      $param->( blast_param){ $tmp1} = $2;
    }
    else {
30    $warning .= "Did not recognize $tmp1 in $_";
    }
    else {
      $warning .= "Not enough values fore blast_param in $_";
    }
35 }
elseif( /^s*(\S+)\s+([^\#]+)/ ) {
  $tmp1 = $1;
  $tmp1 =~ tr/A-Z/a-z/;
  if( defined $param->( $tmp1)) {
40    $tmp2 = $2;
    $tmp2 =~ s/^(.*?)\s*$/ $1/;
    $param->( $tmp1) = $tmp2;
  }
  else {
45    $warning .= "Did not recognize $tmp1 in $_";
  }
  else {
    $warning .= "Key value pair expected, line not recognized
50 $_";
  }
  $warning;
}
55

sub update_derived_param($){
  my ($param) = @_;

```

```

    foreach ( keys %{$param->( oligod_param)}) {
        $param->( oligod_param){ $_}{ squash_factor} = squash_factor(
5          $param->( oligod_param){ $_}{ squash_dx},
          $param->( oligod_param){ $_}{ squash_dy});
    }
}

10 sub initialize_param($){
    my ($popt) = @_;
    my $mat;

    # Parameter priority
15    # 1. Commandline parameters.
    # 2. Parameters from parameter file.
    # 3. Default values.

    my $param = {
20        "fastadb",          "",
        "blastdb",          "/home/nt/database/bf/h_sapiens",
        "oligo_length",      50,
        "oligo_sense",       "direct",
        "cphase",            0,
25        "cfreq",            3,
        "end_spike_len",     0,
        "mindist",           50,
        "maxhits",           8,
        "max_noligo",        10,
30        "min_score", 0, # 35- ignore < 18 matches in oligo
    scoring
        "dnaconc",           2000, # nMol
        "saltconc",          115, # mMol
        "gene_len_cutoff",   50,
35        "gene_ident_cutoff",98,
        "blast_param",       ("wordlen", 11, "strand", "both
    strands",
                                "expect", 50, "nproc", 2, "filter",
    "F"),
40        "lna",              1
    };

    $param->( oligod_param) = {
45        "self_match", { "weight", 1,
                        "cutoff", 50,
                        "squash_dx", 10,
                        "squash_dy", 0.9,
                        "squash_factor", 1,
50                        "layer", 1,
                        },
        "self_hyp", { "weight", 0,
                    "cutoff", 50,
                    "squash_dx", 10,
55                    "squash_dy", 0.9,
                    "squash_factor", 1,
                    "layer", 1,
                    },
    };

```

```

5      "tm_min",      { "weight",      1,
                      "cutoff", 10,
                      "squash_dx",  2,
                      "squash_dy",  0.9,
                      "squash_factor",1,
                      "layer",  0,
                      },
10     "tm_max",      { "weight",      1,
                      "cutoff", 100,
                      "squash_dx",  2,
                      "squash_dy",  0.9,
                      "squash_factor",1,
                      "layer",  0,
                      },
15     "tm",          { "weight",      1,
                      "cutoff", 1,
                      "squash_dx",  0.8,
                      "squash_dy",  0.9,
                      "squash_factor",1,
                      "layer",  1,
                      },
20     "tm_dan_min", { "weight",      1,
                      "cutoff", 10,
                      "squash_dx",  2,
                      "squash_dy",  0.9,
                      "squash_factor",1,
                      "layer",  0,
                      },
25     "tm_dan_max", { "weight",      1,
                      "cutoff", 100,
                      "squash_dx",  2,
                      "squash_dy",  0.9,
                      "squash_factor",1,
                      "layer",  0,
                      },
30     "tm_dan",      { "weight",      1,
                      "cutoff", 1,
                      "squash_dx",  0.8,
                      "squash_dy",  0.9,
                      "squash_factor",1,
                      "layer",  1,
                      },
35     "hit_score",   { "weight",      0,
                      "cutoff", 10000,
                      "squash_dx",  5000,
                      "squash_dy",  0.9,
                      "squash_factor",1,
                      "layer",  1,
                      },
40     "target_struct",{ "weight", 1,
                      "cutoff", 30,
                      "squash_dx",  20,
                      "squash_dy",  0.9,
                      "squash_factor",1,
                      "layer",  1,
                      },
45     "max_match",   { "weight",      1,
                      "cutoff", 30,

```

```

        "squash_dx", 5,
        "squash_dy", 0.9,
        "squash_factor", 1,
        "layer", 1,
5      },
      "max_stretch", { "weight", 1,
        "cutoff", 20,
        "squash_dx", 2,
        "squash_dy", 0.9,
10      "squash_factor", 1,
        "layer", 1,
      },
      "ggg", { "weight", 0,
        "cutoff", 3,
        "squash_dx", 1,
        "squash_dy", 0.9,
15      "squash_factor", 1,
        "layer", 1,
      },
      "cg", { "weight", 0,
        "cutoff", 1,
        "squash_dx", 1,
        "squash_dy", 0.9,
        "squash_factor", 1,
        "layer", 1,
25      },
    };

    if( $popt->{ paramfn } ) {
      local *FILE;
      open( FILE, "< $popt->{ paramfn}" ) ||
        die "Could not open $popt->{ paramfn} $!";
      print read_param( $param, \*FILE);
35    close FILE;
    }

    foreach ( "oligo_length", "oligo_sense", "mindist", "maxhits",
      "max_noligo",
40    "cphase", "cfreq", "end_spike_len", "min_score", "lna",
      "matrix",
      "fastadb", "blastdb", "gene_ident_cutoff",
      "gene_len_cutoff" ) {
      if( defined $popt->{ $_ } ) {
45        $param->{ $_ } = $popt->{ $_ };
      }
      if( defined $popt->{ oligod_param_ggg_weight } ) {
        $param->{ oligod_param }{ ggg }{ weight } = $popt->{
50      oligod_param_ggg_weight };
      }
      if( defined $popt->{ oligod_param_cg_weight } ) {
        $param->{ oligod_param }{ cg }{ weight } = $popt->{
      oligod_param_cg_weight };
55    }
      if( defined $popt->{ oligod_param_tm_min } ) {
        $param->{ oligod_param }{ tm_min }{ cutoff } =
          $popt->{ oligod_param_tm_min };

```

```

    )
    if( defined $popt->( oligod_param_tm_max) ) {
        $param->( oligod_param){ tm_max}{ cutoff} =
            $popt->( oligod_param_tm_max);
5    }
    if( defined $popt->( oligod_param_tm_dan_min) ) {
        $param->( oligod_param){ tm_dan_min}{ cutoff} =
            $popt->( oligod_param_tm_dan_min);
    }
10   if( defined $popt->( oligod_param_tm_dan_max) ) {
        $param->( oligod_param){ tm_dan_max}{ cutoff} =
            $popt->( oligod_param_tm_dan_max);
    }

15   if( $param->( matrix) ) {
        if( ! -e $param->( matrix)) {
            die "Did not find $param->( matrix) $!";
        }
    }
20   else {
        if( $param->( lna) ) { $param->( matrix)      = "$glb->(
dir)/lna.mat"; }
        else { $param->( matrix)      = "$glb->(
dir)/dna.mat"; }
25   if( ! -e $param->( matrix)) {
        if( $param->( lna) ) { $mat = default_mat( "lna"); }
        else { $mat = default_mat( "dna"); }
        write_file( $mat, $param->( matrix));
    }
30   }

    update_derived_param( $param);
    $param;
}

35

sub parse_argv($){
    my ( $glb) = @_;
    my $opt;
40   my $popt;

    GetOptions( "param:s"      => \$popt->( paramfn),
                "conf:s"      => \$glb->( conf),
                "fastadb:s"   => \$popt->( fastadb),
45   "blastdb:s"           => \$popt->( blastdb),
                "length:i"    => \$popt->( oligo_length),
                "sense:i"     => \$popt->( oligo_sense),
                "mindist:i"   => \$popt->( mindist),
                "cphase:i"    => \$popt->( cphase),
50   "cfreq:i"            => \$popt->( cfreq),
                "end_spike_len:i" => \$popt->( end_spike_len),
                "min_tm_dan:i" => \$popt->(
oligod_param_tm_dan_min),
                "max_tm_dan:i" => \$popt->(
55   oligod_param_tm_dan_max),
                "min_tm:i"    => \$popt->( oligod_param_tm_min),
                "max_tm:i"    => \$popt->( oligod_param_tm_max),
                "min_score:i" => \$popt->( min_score),

```

```

        "maxhits:i"      => \${opt}->{ maxhits},
        "maxoligo:i"     => \${opt}->{ max_noligo},
        "lna:i"          => \${opt}->{ lna},
        "gene_ident_cutoff:i" => \${opt}->{
5   gene_ident_cutoff},
        "matrix:s"       => \${opt}->{ matrix},
        "rev"            => \${opt}->{ rev},
        "comp"           => \${opt}->{ comp},
        "revcomp"        => \${opt}->{ revcomp},
10    "capitalize"       => \${opt}->{ capitalize},
        "verbose+"       => \${glb}->{ verbose},
        "view:s"         => \${glb}->{ view},
        ""               => \${opt}->{ stdio}
    );

15    if( \${opt}->{ stdio} ) {
        push @ARGV, "-";
    }

20    if( $#ARGV == -1 ) {
        usage( \${glb}, "No sequence file");
    }

    ( \${opt}, \${popt} );
25 )

sub dump_param( $$ ) {
    my ( \${glb}, \${param} ) = @_;
30    my \${key};
    my \${property};
    my \${value};
    my \${str} = "<PARAMETER>\n";

35    foreach \${key} ( sort keys %\${param} ) {
        if ( \${key} eq "oligod_param" ) {
            foreach \${property} ( sort keys %{\${param}->{ \${key}}} ) {
                foreach \${value} ( sort keys %{\${param}->{ \${key}}{ \${property}}} )
40                {
                    \${str} .= sprintf( "%-16s %-16s %-16s %g\n", \${key}, \${property},
\${value},
                        \${param}->{ \${key}}{ \${property}}{ \${value}} );
                }
            }
45        }
        elsif ( \${key} eq "blast_param" ) {
            foreach \${property} ( sort keys %{\${param}->{ \${key}}} ) {
                \${str} .= sprintf( "%-16s %-16s %s\n", \${key}, \${property},
                    \${param}->{ \${key}}{ \${property}} );
50            }
        }
        else {
            \${str} .= sprintf( "%-16s %s\n", \${key}, \${param}->{ \${key}} );
        }
55    }
    \${str} .= "</PARAMETER>\n";

    \${str};

```

```

    )

##### SEQUENCE I/O #####

5  sub print_fasta($$){
    my ($seq, $filept) = @_;
    my $sq  = $seq->( seq);
    my $str;

10  print $filept ">$seq->( name)";
    if( defined $seq->( comment) ) {
        $str = substr( $seq->( comment), 0, 55);
        print $filept " $str";
15  }
    print $filept "\n";

    if( defined $sq ) {
        $sq  =~ tr/A-Za-z//cd;

20  while( $sq  =~ /(.(60))/gc ) {
            print $filept "$1\n";
        }
        if( $sq  =~ /(.(+)/gc ) {
25  print $filept "$1\n";
        }
    }
    else {
        print $filept "\n";
30  }
}

sub read_fasta($$) {
    my ( $lastline, $filept) = @_;
35  my $seq;

    if( !$lastline) {
        while( <$filept> ) {
            if( /^>/ ) {
40  $lastline = $_;
                last;
            }
        }
    }

45  if( $lastline) {
        if( $lastline =~ />(\S+)\s*(\S*)/ ) {
            $seq->( name)      = $1;
            $seq->( comment)   = $2;
            $seq->( seq)       = "";
50  $lastline                = "";
            while( <$filept> ) {
                if( /^>/ ) {
                    $lastline = $_;
                    $seq->( seq) =~ tr/a-zA-Z//cd;
55  return $seq;
                }
                $seq->( seq)      .= $_;
            }
        }
    }
}

```



```

    }
    else {
        $seq->( name)      = "-";
        $seq->( comment)   = "";
5      $seq->( seq)       = "";
    }
  }
  if( $seq ) {
10    $seq->( seq) =~ tr/a-zA-Z//cd;
  }
  $seq;
}

15 sub print_fasta_fn( $$ ){
    my ( $seq, $fn ) = @_;
    local *FILE;

    open( FILE, ">$fn" ) || die "Could not open $fn $!";
20    print_fasta( $seq, \*FILE );
    close FILE;
}

25 sub write_file( $$ ){
    my ( $str, $fn ) = @_;
    local *FILE;

    open( FILE, ">$fn" ) || die "Could not open $fn $!";
30    print FILE $str;
    close FILE;
}

sub read_file( $ ){
35    my ( $fn ) = @_;
    local *FILE;
    my $str = "";

    open( FILE, "<$fn" ) || die "Could not open $fn $!";
40    while( <FILE> ) {
        $str .= $_;
    }
    close FILE;
    $str;
45 }

##### BLAST #####

50 sub formatblastdb( $$$ ){
    my ( $glb, $dbfn, $type ) = @_;
    my $db = $dbfn;
    my $size;
    my $mtime;
55    my $prev_size = 0;
    my $prev_mtime = 0;
    local *FILE;
    my $d;

```

```

$db = s/(.*)\.(.*)$/1/; # Remove suffix
$db = s/.*\/(.*)/1/; # Remove path

5 print "Do I need to format $db?\n";

($size, $mtime) = stat $dbfn;
if( -e "$glb->{ tmpdir}/$db.stat" ) {
  open( FILE, "<$glb->{ tmpdir}/$db.stat" ) ||
10   die "Could not open $glb->{ tmpdir}/$db.stat $!";
  while( <FILE> ) {
    if( /size=\s*(\d+).*mtime=\s*(\d+)/ ) {
      $prev_size = $1;
      $prev_mtime = $2;
15   }
  }
  close FILE;
}

20 if( ! -e "$glb->{ tmpdir}/$db" || $prev_size != $size ||
    $prev_mtime < $mtime ) {
  print " YES\n";
  $d = "$glb->{ tmpdir}/$db";
  if( -l $d ) {
25   unlink $d;
  }
  if( $dbfn =~ /\^\/ / ) {
    $d = $dbfn;
  }
30 else {
    my $cwd;
    chomp( $cwd = `pwd` );
    $d = "$cwd/$dbfn";
  }
35 print `ln -s $d $glb->{ tmpdir}/$db`;
  print `cd $glb->{ tmpdir}; $glb->{ formatdb} -t $db -i $db -p
$type`;
  print `echo "size= $size mtime= $mtime" > $glb->{
tmpdir}/$db.stat`;
40 }
else {
  print " No, The database is ok.\n";
}
$db = "$glb->{ tmpdir}/$db";
45 $db;
}

sub blast($$$$)(
  my ( $glb, $queryseq, $dbfn, $method, $parameters ) = @_;
50 my $queryfn = "$glb->{ tmpdir}/$.fasta";
  local *FILE;
  local *PIPE;
  my $result = "";

55 open( FILE, ">$queryfn" ) || die "Could not open $queryfn $!";
  print_fasta( $queryseq, \*FILE );
  close FILE;

```

```

# print $gblb->( blastall) $parameters -p $method -d $dbfn -i
$queryfn\n";
open( PIPE,
  qq($gblb->( blastall) $parameters -p $method -d "$dbfn" -i
5  $queryfn {}} ||
  die "Failed to run $gblb->( blastall)";
  while( <PIPE> ) {
    $result .= $_;
10  }
  close PIPE;

  unlink $queryfn;

  $result;
15  }

sub parse_blast($){
  my ($blastout) = @_;
20  my $query = q(Score =\s*(\d+).*?) .
    q(Expect =\s*({\d+E\-.]+).*?) .
    q(\((\d+)%\).*?) .
    q(Strand = (Plus|Minus) \/ (Plus|Minus).*?) .
    q(((Query:\s*(\d+)\s+)([^\s]+)[^>]*) .
25  q( (\d+)[^S]+Sbjct{^>*)\n\n[^>S]*));
  my %result = ();
  my $expect;
  my $name;
  my $score;
30  my $ident;
  my $qstrand;
  my $tstrand;
  my $qstart;
  my $qend;
35  my $tstart;
  my $tend;
  my $alignment;
  my $blastpos1;
  my $blastpos2;
40  my $l1;
  my $s1;
  my $b;
  my $queryseq = "";
  my $matchseq = "";
45  my $targetseq = "";
  my $target;
  my $target_no = 0;
  my $qlen = 0;

50  if( $blastout =~ /^Query=\s*(\S+).*?\((\d+)\s+letters\)/sm ) {
    $result{ query_name} = $1;
    $result{ qlen} = $2;
  }
  else {
55  $result{ query_name} = "-";
    $result{ qlen} = 0;
  }
  $result{ hits} = {};

```

```

while( $blastout =~ /(^(>(\S+)[^>]*)/smgc ) {
    $target      = $1;
    $name        = $2;
5    $target_no++;
    $blastpos1 = pos;
    while( $target =~ /$query/smgc ) {
        #print "$2, $3, $4\n";
        $tstart = '';
10        $score      = $1;
        $expect      = $2;
        $ident       = $3;
        $qstrand     = $4;
        $tstrand     = $5;
15        $alignment  = $6;
        $l           = $7;
        $qstart      = $8;
        $b           = $9;
        $qend        = $10;
20        $blastpos2 = pos;

        $queryseq = "";
        $matchseq = "";
        $targetseq = "";
25        $l       = length( $l );
        $s1       = $l - length("Query:");
        $b       = length( $b );
        while( $alignment =~ /^Query:.(($s1)).(1,$b))\ \d+\n.($l)
            {.(1,$b))\nSbjct:.(($s1)).(1,$b))\ \d+\n/mgcx)
30    {
        $queryseq .= $1;
        $matchseq .= $2;
        $targetseq .= $4;
        if( $tstart eq '' ) {
35            $tstart = $3;
        }
        $tend = $5;
    }

40    #print "\n1: $queryseq\n2: $matchseq\n3: $targetseq\n";

    $expect =~ s/^e/le/;
    $qlen = $qend - $qstart + 1;
    push @{$result( hits)},
45    ( "tname", $name, "score", $score, "expect", $expect,
      "ident", $ident,
      "qstrand", $qstrand, "tstrand", $tstrand, "target_no",
      $target_no,
      "qstart", $qstart, "qend", $qend, "qlen", $qlen,
50      "tstart", $tstart, "tend", $tend,
      "self", 0,
      "queryseq", $queryseq, "matchseq", $matchseq,
      "targetseq", $targetseq);
    pos      = $blastpos2;
55    }
    pos      = $blastpos1;
}
$result;

```

```

)

##### perceptron #####
5
sub squash_factor($$){
  # This factor determines how steep the squashing function
  operates
  # a change of $x from 0 gives a $y response if used with $k
10  my ($x, $y) = @_;
    my $k = - log(1/$y - 1) / $x;
    $k;
  }

15 sub squash($){
    my ($x) = @_;
    my $y = 1/(1+exp(-$x));
    $y;
  }

20 sub dot($$){
    my ($a, $b) = @_;
    my $sum = 0;
    my $i;

25    if( $$a != $$b){
      die "dot: a ($$a) and b ($$b) should have the same number of
      elements";
    }

30    for( $i=0; $i<=$$a; $i++) {
      #if( ! defined $a->[ $i] ) { die "OH OH a $i was not defined";
    }
      #if( ! defined $b->[ $i] ) { die "OH OH b $i was not defined";
35    }
      $sum += $a->[ $i] * $b->[ $i];
    }

    $sum;
40  }

sub sum($){
  my ($a) = @_;
  my $sum = 0;
45  my $i;

  for( $i=0; $i<=$$a; $i++) {
    $sum += $a->[ $i];
  }

50  $sum;
  }

sub perceptron($$){
55  my ($weight, $input) = @_;
    my $sum = sum( $weight);
    #my $output = squash( dot( $weight, $input));
    my $output = dot( $weight, $input)/ $sum;

```

```

    $output;
}

5
##### SYSTEM #####

sub openfile($) {
    my ($fn) = @_;
10    local *FILE;
    open( FILE, $fn ) ||
        die "Could not open $fn $!\n";

    *FILE;
15 }

sub mkdir($) {
    my ($dir) = @_;
20    my @dir = split(/\//, $dir);
    my $d = '';

    foreach (@dir) {
        $d .= $_;
25        if( $d && ! -e $d ) { mkdir $d; }
        $d .= "/";
    }
}

30
sub usage($$) {
    my ($glb, $message) = @_;
    my $file = openfile( $glb->{ program_name} );

35    $message .= "\n";
    while (<$file) {
        if( /^#(.*)/ ) {
            $message .= "$1\n";
        }
40    else {
        --last;
    }
    }
    close $file;
45    die "$message";
}

sub min(@) {
    my $min = 9999999999;
50    foreach (@_) { if( $_ < $min ) { $min = $_; } }
    $min;
}

sub max(@) {
55    my $max = -9999999999;
    foreach (@_) { if( $_ > $max ) { $max = $_; } }
    $max;
}

```

```

sub up($) {
  my ( $str ) = @_;
  $str =~ tr/a-z/A-Z/;
5  $str;
}

##### SEQUENCE FUNCTIONS #####

10 sub comp($) {
  my ( $seq ) = @_;
  $seq =~
tr/acgtumrwsykvhdbxnACGTUMRWSYKVHDBXN/tgcaakywsrmbdhvxxTGCAAKYWS
15 RMBDHVXX/;

  $seq;
}

20 sub rev($) {
  my ( $seq ) = @_;
  $seq = reverse( split( " ", $seq ) );

  $seq;
25 }

sub revcomp($) {
  my ( $seq ) = @_;
  $seq = rev( $seq );
30 $seq = comp( $seq );
  $seq;
}

sub capitalize($$$$){
35 my ( $seq, $phase, $freq, $end_spike_len ) = @_;
  my $pat = "." x $phase;
  my $n = $freq - $phase - 1;

  $seq =~ tr/A-Z/a-z/;
40
  if( $freq != 0 ) {
    if( $phase < 0 || $phase >= $freq ) {
      die "phase ($phase) is out of range\n";
    }
45    if( $freq < 1 ) {
      die "freq ($freq) is out of range\n";
    }

    $seq =~ s/($pat){.}{(0,$n)}/$1 . up($2) . $3/eg;
50 }
    if( $end_spike_len != 0 ) {
      $seq =~ s/((0,$end_spike_len))(.*)(($end_spike_len))/up($1)
. $2 . up($3)/eg;
55 }

  $seq;
}

```

```

##### MELTING TEMPERATURE #####

sub melt_temp($) {
  my ($seq) = @_;
5   my $temp;
   my $n_at;
   my $n_gc;
   my $n_ot;

10   $seq    =~ tr/a-zA-Z//cd;

   $_ = $seq;
   $n_at = tr/aAtT//d;
   $_ = $seq;
15   $n_gc = tr/gGcC//d;
   $_ = $seq;
   $n_ot = tr/aAtTgGcC//cd;

   $temp = (0.88 * $n_at) + (1.47 * $n_gc) + $n_ot;

20   ($n_at, $n_gc, $n_ot, $temp);
}

sub tm_pred($) {
25   my ($seq) = @_;
   my $temp = 0;
   my $conf = 0;

30   if(1){
     $seq =~ tr/acgtACGT//cd;

     #Oligoconc is the sum of target and probe concentrations molar
35

     my $output =
       `$glb->{ lynx} -cookie_file=$glb->{ lynx_cookie} -source $glb-
       >{
40   tm_exiqon_url}?sequences='$seq'&saltconc='0.115'&oligoconc='0.00
       -0002';

       print "$glb->{ lynx} -cookie_file=$glb->{ lynx_cookie} -source
       $glb->{ tm_exiqon_url}?sequences='$seq'\n";
45   print "$output\n";

       if( $output =~
       /<td>([\d.]+)</td>\s+<td>([\d.]+)</td></tr>/s) {
         $temp = $1;
50         $conf = $2;
       }
     }

     ($conf, $temp);
55   }

sub tm($$@) {

```



```

my ($glb, $param, @seq) = @_;
my $temp = 99999;
my $conf = 0;
my $comand;
5 my $seqfn = "$glb->( tmpdir)/${$}tm.seq";
my $resultfn = "$glb->( tmpdir)/${$}tm.dat";
local *FILE;
local *OLDERR;
my $score_min;
10 my $score_max;
my $score;
my $wsum;
my $req;
my $res;
15 my $output;
my @result;
my $seq;
my $content = '';

20 foreach $seq ( @seq) {
    $seq =~ tr/acgtACGT//cd;
    if( $content ) {
        $content .= "\n$seq";
    }
    else {
        $content = "sequences=$seq";
    }
}

30 # $seq = $seq[0];

    $req = HTTP::Request->new( POST => $glb->( tm_url));
    $req->content_type('application/x-www-form-urlencoded');
35 # $req->content('tmuser=oligod&email=tolstrup@exiqon.com');
    # $req->content("sequences=$seq");
    $req->content( $content);

    $res = $glb->( ua)->request($req);
40
    #print "$seq\n";
    # $output = $res->content;
    #print "$output\n";

45 # check the outcome
    if ($res->is_success) {
        $output = $res->content;
    } else {
        print "Unsuspected Tm result:" . $res->status_line . "\n";
50 }

    #my $output =
    # ` $glb->( lynx) -cookie_file=$glb->( lynx_cookie) -source
55 $glb->( tm_exiqon_url)?sequences='$seq';

    # print "#### THE OUTPUT WAS #####\n$output\n";

```

```

if( defined $output ) (
  while( $output =~
/<td>([\d.]+)<\/td>\s+<td>([\d.]+)<\/td><\/tr>/gis) {
  $temp      = $2;
5
  $score_min = squash( $param->( oligod_param){ tm_min}(
squash_factor) *
    ($temp - $param->( oligod_param){ tm_min}( cutoff)));
  $score_max = squash( $param->( oligod_param){ tm_max}(
10 squash_factor) *
    ($param->( oligod_param){ tm_max}( cutoff) - $temp));

  $wsum      = $score_min * $param->( oligod_param){ tm_min}(
weight) +
15 $score_max * $param->( oligod_param){ tm_max}(
weight);

  $score     = squash( $param->( oligod_param){ tm}(
squash_factor) *
20 ($wsum - $param->( oligod_param){ tm}( cutoff)));

  push @result, ( "tm",      ( "raw", $wsum, "score", $score),
    "tm_min", ( "raw", $temp, "score", $score_min),
    "tm_max", ( "raw", $temp, "score", $score_max));
25 )
}
else {
  print STDERR "Sorry, the tm prediction is not available.";
  foreach (@seq) {
30   push @result, ( "tm",      ( "raw", 0, "score", 0),
    "tm_min", ( "raw", 0, "score", 0),
    "tm_max", ( "raw", 0, "score", 0));
  }
35 }

\@result;
}

40
sub tm_dan($$$){
  my ($glb, $param, $seq) = @_;
  my $temp = 9;
  my $comand;
45 my $seqfn = "$glb->( tmpdir)/${$}dan.seq";
  my $resultfn = "$glb->( tmpdir)/${$}dan.dat";
  local *FILE;
  local *OLDERR;
  $seq =~ tr/acgtACGT//cd;
50 my $windowsize = length( $seq);
  my $score_min;
  my $score_max;
  my $score;
  my $wsum;
55

  write_file( $seq, $seqfn);

  $comand = "$glb->( dan) -sequence $seqfn -windowsize ".
```

```

        "$windowsize -shiftincrement 1 ".
        "-dnaconc $param->( dnaconc) -saltconc $param->(
saltconc) ".
        "-outfile $resultfn";
5
    open(OLDERR, ">&STDERR");
    open(STDERR, '>', "/dev/null");
    my $output = `scomand`;
    close(STDERR);
10    open(STDERR, ">&OLDERR");

    open( FILE, "<$resultfn" ) || die "Could not open $resultfn";
    while( <FILE> ) {
        if( /Tm=({\d.}+)/ ) {
15            $temp = $1;
        }
        elsif( /^s\d+\s\d+\s+([0-9.]+)/ ) {
            $temp = $1;
20        }
    }
    close FILE;

    unlink $seqfn;
    unlink $resultfn;
25

    $score_min = squash( $param->( oligod_param){ tm_dan_min}{
squash_factor} *
        ($temp - $param->( oligod_param){ tm_dan_min){
30    cutoff}));
    $score_max = squash( $param->( oligod_param){ tm_dan_max}{
squash_factor} *
        ($param->( oligod_param){ tm_dan_max){ cutoff} -
35    $temp));
    $wsum = $score_min * $param->( oligod_param){ tm_dan_min}{
weight} +
        $score_max * $param->( oligod_param){ tm_dan_max){ weight};

40    $score = squash( $param->( oligod_param){ tm_dan){
squash_factor} *
        ( $wsum - $param->( oligod_param){ tm_dan){ cutoff}));

45    ( ( "raw", $wsum, "score", $score),
      ( "raw", $temp, "score", $score_min),
      ( "raw", $temp, "score", $score_max));
    )

##### SELF HYBRIDISATION #####
50
sub default_mat($){
    my ( $mat_name) = @_;
    my $mat;

55    if( $mat_name eq "lna" ) {
        $mat =
";D LNA hybridisation scoring matrix
1 45 80 5 6 80 4

```

```

-8 -50
* @ 0 1 2
ACGTRYMWSKDHVBNLIOU
0 1 2 3 0 1 0 0 1 2 0 0 0 1 0 0 1 2 3
5  -4
   -4 -4
   -4 5 -4
   3 -4 -4 -4
   1 -1 1 -1 1
10 -1 1 -1 1 -4 1
    1 1 -2 -2 -1 -1 1
    1 -2 -2 1 -1 -1 -1 1
    -2 1 1 -2 -1 -1 -1 -1 1
    -2 -2 1 1 -1 -1 -1 -1 -1 1
15 1 -2 1 1 1 -1 -1 1 -1 1 1
    1 1 -2 1 -1 1 1 1 -1 -1 -1 1
    1 1 1 -2 1 -1 1 -1 1 -1 -1 -1 1
    -2 1 1 1 -1 1 -1 -1 1 1 -1 -1 -1 1
    1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 1
20 -4 -4 -4 6 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
    -4 -4 6 -4 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
    -4 6 -4 -4 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 8 -1
    6 -4 -4 -4 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 8 -1 -1
";
}
25  elseif( $mat_name eq "dna" ) {
      $mat =
";D hybridisation scoring matrix
1 45 80 5 6 80 4
30  -8 -50
    * @ 0 1 2
    ACGTRYMWSKDHVBN
    0 1 2 3 0 1 0 0 1 2 0 0 0 1 0
    -4
35  -4 -4
    -4 5 -4
    3 -4 -4 -4
    1 -1 1 -1 1
    -1 1 -1 1 -4 1
40  1 1 -2 -2 -1 -1 1
    1 -2 -2 1 -1 -1 -1 1
    -2 1 1 -2 -1 -1 -1 -1 1
    -2 -2 1 1 -1 -1 -1 -1 1
    1 -2 1 1 1 -1 -1 1 -1 1 1
45  1 1 -2 1 -1 1 1 1 -1 -1 -1 1
    1 1 1 -2 1 -1 1 -1 1 -1 -1 1
    -2 1 1 1 -1 1 -1 -1 1 1 -1 -1 -1 1
    1 1 1 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 1
";
50  }
    else{
      die "Ups $mat_name was not recognised try lna or dna";
    }
55  $mat;
}

```

```

sub self_match($$$){
  my ( $glb, $param, $seq) = @_;
  local *PIPE;
5   my $seqfn    = "$glb->( tmpdir)/${$}seq.fasta";
  my $compseqfn = "$glb->( tmpdir)/${$}compseq.fasta";
  my $f_seq;
  my $score      = 0;
  my $norm_score = 0;
10  my $alignment = "";
  local *OLDERR;
  my $evidence = '';
  my $ssparam  = '';
  my $str;
15
  if( $param->( matrix) ) {
    $ssparam = " -s $param->( matrix) ";
  }

20  open(OLDERR, ">&STDERR");
  open(STDERR, ">/dev/null") || die "Can't open /dev/null
  $!";

  $f_seq->( seq) = $seq;
  $f_seq->( name) = "seq";
25  if( $param->( lna) ) {
    $f_seq->( seq) =~ tr/ACGT/LIOU/;
  }

30  print_fasta_fn( $f_seq, $seqfn);
  $f_seq->( seq) = rev( $seq);
  $f_seq->( name) = "rev";

  if( $param->( lna) ) {
35    $f_seq->( seq) =~ tr/ACGT/LIOU/;
  }
  print_fasta_fn( $f_seq, $compseqfn);

  if( $glb->( verbose) > 1 ) { print "<SSEARCH>"; }
40  #print "$glb->( ssearch) $ssparam $seqfn $compseqfn\n";

  open( PIPE, "$glb->( ssearch) $ssparam $seqfn $compseqfn |" ) ||
    die "Failed to open pipe $!";
45  while( <PIPE){
    if( /Smith-Waterman score:\s*(\d+)/ ) {
      $score = $1;
    }
    if( $glb->( verbose) > 1 ) { print; }
50    if( /^[\.:\\t]+$/) {
      tr/./: /;
    }
    if( $param->( lna) ) {
      if( /^(\\S+)(.*)/) {
55        $alignment .= $1;
        $str        = $2;
        $str        =~ tr/ACGTLIOU/acgtACGT/;
        $alignment .= "$str\n";
      }
    }
  }
}

```

```

    }
    else (
        $alignment .= $_;
    )
5   }
    else (
        $alignment .= $_;
    )
}
10 close PIPE;
    if( $glb->( verbose) > 1 ) { print "</SSEARCH>\n"; }

    if( $alignment =~ /(Smith-Waterman score.*)Library sca/s){
        $evidence = $1;
15   if( $glb->( verbose) > 1 ) {
        print "<SELF_EVIDENCE>\n";
        print "$evidence";
        print "</SELF_EVIDENCE>\n";
20   }
    }

    close(STDERR);
    open(STDERR, ">&OLDERR");
25

    unlink $seqfn;
    unlink $compseqfn;

    $norm_score = squash( $param->( oligod_param)( self_match)(
30   squash_factor) *
        ( $param->( oligod_param)( self_match)( cutoff) -
        $score) );

    ( "score", $norm_score, "raw", $score, "evidence", $evidence);
35 }

sub self_hyp($$@){
    my ( $glb, $param, @seqs) = @_;
40   local *PIPE;
    my $score = 0;
    my $norm_score = 0;
    my $evidence = '';
    my $seqfn = "$glb->( tmpdir)/${$}seq_d.fasta";
45   local *FILE;
    my $seq;
    my $i = 0;
    my @result = ();
    my $dyp_opt = '-min_score 0 -max_hyp 1';
50   my $read;

    open( FILE, "> $seqfn" ) || die "Could not open $seqfn";
    foreach ( @seqs ) {
55       $seq->( seq) = $_;
        $seq->( name) = sprintf( "seq_%03d", $i);
        $i++;
        print_fasta( $seq, \*FILE);
    }
}

```

```

    }
    close FILE;

    if( $glb->( verbose) > 1 ) { print "<DYP>"; }
5    #print "$glb->( dyp) -f $seqfn $dyp_opt\n";

    open( PIPE, "$glb->( dyp) -f $seqfn $dyp_opt|" ) || die "Failed
to open pipe $!";
10    $read = 1;
    while( <PIPE>){
        if( $glb->( verbose) > 1 ) { print; }
        if( />(\S*)/ ) {
            }
15        elsif( /Score=\s*(\d+)/ ) {
            if( $evidence ) {
                $norm_score = squash( $param->( oligod_param){ self_hyp}{
squash_factor} *
                ( $param->( oligod_param){ self_hyp){ cutoff} -
20        $score) );
                my $evidence_str = "$evidence";
                push @result,
                ( "score", $norm_score, "raw", $score, "evidence",
$evidence_str);
25                $evidence = '';
            }
            $score = $1;
            $read = 1;
        }
        elsif( /Mask:/ ) {
30            $read = 0;
        }
        elsif( $read ) {
            $evidence .= $_;
35        }
    }
    close PIPE;

    $norm_score = squash( $param->( oligod_param){ self_hyp){
squash_factor} *
40    ( $param->( oligod_param){ self_hyp){ cutoff} -
    $score) );
    push @result, ( "score", $norm_score, "raw", $score,
    "evidence", $evidence);
45

    if( $glb->( verbose) > 1 ) { print "</DYP>\n"; }

    if( $glb->( verbose) > 1 ) {
        print "<SELF_EVIDENCE>\n";
50        print "$evidence";
        print "</SELF_EVIDENCE>\n";
    }

    @result;
55 }

sub target_struct($$$){

```

```

my ( $glb, $param, $seq) = @_;
local *PIPE;
my $score = 0;
my $evidence = '';
5 my $seqfn = "$glb->( tmpdir)/${$}seq_t.fasta";
my @result = ();
my $min_score = 15;
my $dyp_opt = "-depth 100 -min_score $min_score";
my $read_mask = 0;
10 my $mask;

    print_fasta_fn( $seq, $seqfn);

15 #print "$glb->( dyp) -f $seqfn $dyp_opt";

    open( PIPE, "$glb->( dyp) -f $seqfn $dyp_opt|" ) || die "Failed
to open pipe $!";
    while( <PIPE> ){
20 print;
        if( $glb->( verbose) > 1 ) { print; }
        if( />(\S*)/ ) {
            $mask = '';
        }
        elsif( /Score=\s*(\d+)/ ) {
25             $score = $1;
        }
        elsif( /Mask:/ ){
            $read_mask = 1;
30         }
        elsif( $read_mask ) {
            $mask .= $_;
        }
        else {
35             $evidence .= $_;
        }
        close PIPE;

40     print "<TARGET_STRUCT>\n";
    print $evidence;
    if( $mask ) {
        print $mask;
    }
45     else {
        print "No significant secondary structure identified.";
    }
    print "</TARGET_STRUCT>\n";

50     $evidence =~ s/\n//g;
    $mask =~ s/\n//g;

    if( $score < $min_score ) {
55         $mask = "";
    }

    $mask;
}

```



```

##### PATTERN FILTERS #####

5  sub target_struct_score($$$$){
    my ($mask, $start, $end, $param) = @_;
    my $score = 1;
    my $count = 0;

10  $_ = "";

    if( $mask ) {
        my $len = $end - $start;
15    $_ = substr( $mask, $start, $len);
        $count = tr/-/#/;
        my $match = int( $count / $len * 100);
        $score = squash( $param->{ oligod_param}{ target_struct}{
squash_factor} *
20    ( $param->{ oligod_param}{ target_struct}{ cutoff} -
$match) );
    }

    ( "score", $score, "raw", $count, "evidence", $_);
25 }

sub cg_score($){
    my ($seq) = @_;
    if( $seq =~ /cg/i) {
30    return { "score", 0, "raw", 1};
    }
    { "score", 1, "raw", 0};
}

35 sub ggg_score($$){
    my ($param, $seq) = @_;
    my $maxg = 0;
    my $ng;
    my $score;

40  while( $seq =~ /(g+)/gi) {
        $ng = length( $1);
        if( $ng > $maxg) { $maxg = $ng; }
    }

45  $score = squash( $param->{ oligod_param}{ ggg}{ squash_factor}
    *
        ( $param->{ oligod_param}{ ggg}{ cutoff} - $maxg) );
    ( "score", $score, "raw", $maxg);
50 }

##### PROBE #####

55 sub oligo_property_score($$){
    my ($param, $oligo_property) = @_;
    my $oligo_par;

```

```

my $name;
my @input;
my @weight;
my $output;

5  foreach $name ( keys %{$param->( oligod_param)} ) {
    $oligo_par = $param->( oligod_param)( $name);
    if( $oligo_par->( weight) && $oligo_par->( layer) ) {
        #print "name    = $name\n";
10     #print "score    = $oligo_property->( $name){ score}\n";
        #print "weight   = $oligo_par->( weight)\n";
        push @input, $oligo_property->( $name){ score};
        push @weight, $oligo_par->( weight);
    }
15 }
    $output = perceptron( \@weight, \@input);
    $output;
}

20 sub print_oligo_property($$)(
    my ($param, $oligo_property) = @_;
    my $name;
    my $oligo_par;

25     print "<SCORE_EVIDENCE>\n";
    printf( "%-11s %5.2f\n", "oligo score :", $oligo_property->(
score));
    foreach $name ( sort keys %{$param->( oligod_param)} ) {
30     $oligo_par = $param->( oligod_param)( $name);
        if( $oligo_par->( weight)) {
            printf( " %-13s %7g score=%5.2f x %3g (%6g %5g %4g)\n",
$name,
35             $oligo_property->( $name){ raw},
            $oligo_property->( $name){ score},
            $oligo_par->( weight),
            $oligo_par->( cutoff),
            $oligo_par->( squash_dx),
            $oligo_par->( squash_dy));
40     }
    }
    print "</SCORE_EVIDENCE>\n\n";
}

45 sub collect_score($$$$$$)(
    my ($prev_region,
        $start_count, $end_count, $start_score, $end_score,
    $prev_pos,
50     $region) = @_;
    my $next_region;
    my $extra_region;

    if( $start_count && $end_count ) {
55     if( $prev_region ) {
        $prev_region->( end)    = $prev_pos - 1;
        $next_region->( hit_count) = $prev_region->( hit_count);
        $next_region->( score) = $prev_region->( score);
    }

```

```

    $extra_region->( hit_count)= $prev_region->( hit_count);
    $extra_region->( score)    = $prev_region->( score);
}
5  $next_region->( start) = $prev_pos + 1;
    $extra_region->( start)= $prev_pos;
    $extra_region->( end)   = $prev_pos;
    $extra_region->( length)    = 1;
    $extra_region->( hit_count)+= $start_count;
10  $extra_region->( score) += $start_score;

    $next_region->( start) = $prev_pos + 1;
    $next_region->( hit_count) += $start_count;
    $next_region->( score) += $start_score;
15  $next_region->( hit_count) -= $end_count;
    $next_region->( score) -= $end_score;
}
elseif( $start_count){
20  if( $prev_region ) {
    $prev_region->( end) = $prev_pos - 1;
    $next_region->( hit_count)= $prev_region->( hit_count);
    $next_region->( score)= $prev_region->( score);
}
    $next_region->( start) = $prev_pos;
25  $next_region->( hit_count) += $start_count;
    $next_region->( score) += $start_score;
}
elseif( $end_count){
30  if( $prev_region ) {
    $prev_region->( end) = $prev_pos;
    $next_region->( hit_count)= $prev_region->( hit_count);
    $next_region->( score)= $prev_region->( score);
}
    $next_region->( start) = $prev_pos + 1;
35  $next_region->( hit_count) -= $end_count;
    $next_region->( score) -= $end_score;
}
else {
40  $next_region->( start) = $prev_pos;
    $next_region->( hit_count) = 0;
    $next_region->( score) = 0;
}

    if( $prev_region) {
45  $prev_region->( length) = $prev_region->( end) -
    $prev_region->( start)+1;
    if( $prev_region->( length) >0 ) {
        push @$region, $prev_region;
    }
50  }
    if( $extra_region ) {
        push @$region, $extra_region;
    }
    ($prev_region, $next_region);
55  }

sub print_hit($$){
    my ($hit, $newline) = @_;

```

```

my $n1 = " ";
if( $newline) { $n1 = "\n";}

printf( "%4d - %4d %3d%% %4d %10s %6d - %6d %s $n1",
5      $hit->( qstart), $hit->( qend),
      $hit->( ident), $hit->( score),
      $hit->( tname), $hit->( tstart), $hit->( tend),
      $hit->( tstrand));
}
10 ##### IDENTIFY THE SEARCH SEQUENCE IN THE DATABASE #####

sub check_self_hit($$){
15   my ($qlen, $hit) = @_;
   my $glb_ident = 0;
   if( $qlen ) {
       $glb_ident = (($hit->( qend) - $hit->( qstart)) * $hit->(
20   ident))/ $qlen ;
   }
   if( $glb_ident > 95) {
       return $glb_ident;
   }
   0;
25 }

sub overlap($$){
   my ($hit1, $hit2) = @_;
30   my $overlap = 1;
   my $minslack = 5;
   my $maxslack = 30;
   my $len = min( $hit2->(qend)-$hit2->( qstart),
                 $hit1->(qend)-$hit1->( qstart));
35   my $slack = max( $minslack, min( $maxslack, $len * 0.1));
   if( ( $hit1->( qstart) + $slack) > $hit2->(qend) ||
       ( $hit2->( qstart) + $slack) > $hit1->(qend) ) {
       $overlap = 0;
   }
40   $overlap;
}

sub check_coverage($@){
45   my ($qlen, @exon) = @_;
   my $hit;
   my $prev_end = 0;
   my $overlap;
   my $maxgap = 30;
50   my $maxoverlap = 0;

   print "<GENE>\n";
   print "Coverage of the query ($qlen):\n";
   foreach $hit ( sort ($a->( qstart) <=> $b->( qstart)) @exon) {
55       $overlap = $hit->( qstart) - $prev_end;
       if( abs( $overlap) > $maxoverlap) { $maxoverlap = abs(
$overlap); }
       if( $prev_end) {

```

```

        print "$overlap\n";
    }
    print_hit( $hit, 0);
    $prev_end = $hit->{ qend};
5   }
    if( $prev_end) {
        $overlap = $qlen - $prev_end;
        if( abs( $overlap) > $maxoverlap) { $maxoverlap = abs(
10   $overlap); }
        print "$overlap\n";
    }
    if( $maxoverlap > $maxgap ) {
        print "Warning: The query sequence was not found completely in
15   ".
        "the database\n";
    }
}

20 sub find_self($%){
    my ( $param, %blast_parse) = @_;
    my $hit;
    my @exon;
    my $overlap_flag;
25   my $overlap_hit;

    foreach $hit ( sort ($b->{score} <=> $a->{score}))
    @($blast_parse{ hits})) {
        if( $hit->{ qlen} >= $param->{ gene_len_cutoff} &&
30         $hit->{ ident} >= $param->{ gene_ident_cutoff} ) {
            $hit->{ self} = 1;
            $overlap_flag = 0;
            foreach (@exon) {
                if( overlap( $_, $hit) ) {
35                 $overlap_flag = 1;
                my $new_var = $_;
                $overlap_hit = \"$new_var;
                last;
            }
40         }
        if( !$overlap_flag)- {
            push @exon , $hit;
        }
        else {
45             #print "Overlap to:\n";
            #print_hit( $overlap_hit, 1);
        }
    }
}

50 check_coverage( $blast_parse{ qlen}, @exon);

}

55 sub print_exon(@){
    my (@exon) = @_;
    my $hit;
    my $i = 0;

```

```

    foreach $hit ( sort ( $a->( qstart) <=> $b->( qstart)) @exon) (
        $i++;
        print "$hit->( qstart) - $hit->( qend)\n";
    )
5  )

#####

10
sub blast_param2param_line($) {
    my( $blast_param) = @_;
    my %key2opt = ( "wordlen", "-W", "strand", "-S", "expect", "-e",
15         "filter", "-F", "nproc", "-a");
    my %strand2opt = ( "-", 3, "both strands", 3,
        "direct strand only", 1, "reverse strand only", 2,
        "top", 1, "bottom", 2);
    my $param_line = '';
20
    foreach (keys %$blast_param) {
        if( $key2opt{ $_} ) {
            if( $_ eq "strand" ) {
                $param_line .= " $key2opt{ $_} $strand2opt{
25 $blast_param->{ $_}}";
            }
            else {
                $param_line .= " $key2opt{ $_} $blast_param->{ $_}";
30         }
            }
        }
    }
    $param_line;
}

35
sub param_line2blast_param($) {
    my( $param_line) = @_;
    my %key2opt = ( "wordlen", "-W", "strand", "-S", "expect", "-e",
40         "filter", "-F", "nproc", "-a");

    my %opt2strand = ( 3, "both strands", 1, "direct strand only",
        2, "reverse strand only");
    my $blast_param;
45
    my %opt2key = reverse %key2opt;
    my %param_hash = split( " ", $param_line);

    foreach (keys %param_hash) {
        if( $opt2key{ $_} ) {
50             if( $_ eq "-S" ) {
                $blast_param->{ $opt2key{ $_} } = $opt2strand{
                    $param_hash{ $_} };
            }
            else {
55                 $blast_param->{ $opt2key{ $_} } = $param_hash{ $_};
            }
        }
    }
    else {

```

```

        die "Sorry, the option $_ has not been implemented yet.";
    }
    }
    $blast_param;
5  }

sub position_score($$$){
    my ( $glb, $param, $seq) = @_;
10    my $hit;
    my $hit_id = 0;
    my $blast_out;
    my %blast_parse;
    # -e 100 : show all hits 50 = down to 17 nuc in human
15    # -F F      : Don't filter
    # -S 1      : Search only top strand of query, default is
3=both strands
    # -a 4      : Use four CPU's
20    # -W 30    : Word lengt 30 (default is 11)

    my @end_points;
    my @region;
    my $prev_pos = 1;
    my $p;
25    my $score;
    my $hit_count;
    my $prev_region;
    my $next_region;
    my $one_nuc_region;
30    my $start_score;
    my $end_score;
    my $start_count;
    my $end_count;
    my $ident;
35    my $ignore = 0;
    my $param_line = blast_param2param_line( $param->{
blast_param});

    $blast_out = blast( $glb, $seq, $param->{ blastdb}, "blastn",
40    $param_line);
    if( $glb->{ view} =~ /BLAST/ ) {
        print "<BLAST>$blast_out</BLAST>\n";
    }
    %blast_parse= parse_blast( $blast_out);
45    find_self( $param, %blast_parse);
    # push @($blast_parse{ hits}), {
    "query_name","X","tname","X","score",10,
    #
    "qstrand","P","tstrand","P","qstart",252,"qend",269,"ident",
50    0);
    foreach $hit (@($blast_parse{ hits})) {
        #if( $ident = check_self_hit( $blast_parse{ qlen), $hit)) {
        if( $hit->{ self}) {
            if( ! $ignore ) {
55                $ignore = 1;
                print "Ignoring the following hit(s):\n";
            }
            $ident = sprintf( "%3.0f", $hit->{ ident});

```

```

    printf( "%30s %6d %6d %s\n", $hit->( tname), $hit->(
tstart), $hit->( tend),
    "$hit->( ident)% identity to $blast_parse( query_name)");
}
5   else(
    push @end_points, ("hit_id", $hit_id, "pos", $hit->(qstart),
"type","start");
    push @end_points, ("hit_id", $hit_id, "pos", $hit->(qend),
10  "type","end");
  if(0){
    print "$blast_parse( query_name), $hit->( tname) ".
      "$hit->(score) ".
      "$hit->(expect) ".
15    "$hit->(qstrand) ".
      "$hit->(tstrand) ".
      "$hit->(qstart) ".
      "$hit->(qend) ".
      "$hit->(ident)\n";
20  }
    $hit_id++;
  }
}

25  if( ! $ignore ) {
    print "The query sequence was not found in the database!\n";
  }
  print "</GENE>\n";

30  $start_count      = 0;
  $end_count         = 0;
  $start_score       = 0;
  $end_score         = 0;

35  if( $#end_points >= 0 ) {

    foreach (sort { $a->(pos) <=> $b->(pos)} @end_points) {
      # print "$_->( pos)\n";

40    if( $_->( pos) > $prev_pos) {
      $prev_region = collect_score( $prev_region,
      $start_count, $end_count, $start_score, $end_score,
      $prev_pos,
      \@region);
45    $prev_pos      = $_->( pos);
      $start_count = 0;
      $end_count   = 0;
      $start_score = 0;
      $end_score   = 0;
50    }

    if( $_->( type) eq "start" ) {
      $start_score += $blast_parse( hits)->[ $_->( hit_id)]->(
score);
55    $start_count ++;
    }
    else {

```



```

        $end_score += $blast_parse( hits)->[ $_->( hit_id)]->(
score);
        $end_count ++;
    }
5    }

    ($prev_region, $p) = collect_score( $prev_region,
    $start_count, $end_count, $start_score, $end_score,
    $prev_pos,
10    \@region);
    my $last_region;
    $last_region->( start) = $prev_region->( end) + 1;
    $last_region->( end) = length( $seq->( seq));
    $last_region->( length) =
15    $last_region->( end) - $last_region->( start) +
    1;
    $last_region->( score) = 0;
    $last_region->( hit_count) = 0;
    if( $last_region->( length) > 0 ) {
20    push @region, $last_region;
    }
}

25 if(0){
    foreach (@region) {
        print
        "$_->( start), $_->( end), $_->( score), $_->( hit_count),
        $_->( length)\n";
30    }
}

    (\@region, \%blast_parse);
35 }

sub hit_score($$$$){
    my ($position_score, $start, $end, $param) = @_;
    my $sumscore = 0;
40    my $n = 0;
    my $score;
    my $overlap;
    my $maxscore = -9999999999;
    my $norm_score;
45    my $result;

    foreach (@$position_score) {
        if( ( $_->( start) <= $end && $_->( end) >= $start) ) {
            $overlap = min( $end, $_->( end)) - max( $start, $_->(
50    start));
            $score = $_->( score) * $overlap; # / $_->( length);
            if( $score > $maxscore ) {
                $maxscore = $score;
            }
55    if( $score > $param->( min_score) ) {
            $sumscore += $score;
            #print "($score $_->( start) $_->( end) $_->( score) $_->(
            length) $overlap)";

```

```

    )
    $n ++;
  }
}
5  $norm_score = squash( $param->{ oligod_param}{ hit_score){
  squash_factor} *
  ( $param->{ oligod_param}{ hit_score}{ cutoff} -
  $sumscore) );
  #print "SCORE: $sumscore, $param->{ min_score}, $norm_score\n";
10  $$score = int( $sumscore + 0.5); # 0.5 to get correct rounding
  $result = ( "score", $norm_score, "raw", $sumscore);

  $result;
15  )

sub oligo_sort_diff($$){
  my ($a,$b) = @_;
  my $result;
20  $result = $a->{dbmatch} <=> $b->{ dbmatch};
  if( ! $result) { $result = $a->{temp} <=> $b->{ temp}; }
  #if( ! $result) { $result = $a->{self_score} <=> $b->{
  self_score}; }
  $result;
25  )

sub find_oligo_hits($$$){
  my ($blast_parse, $start, $end) = @_;
  my $hit;
  my $overlap;
30  my $qoffset;
  my @hits = ();
  my $qseq;
  my $mseq;
  my $tseq;
35  my $l = '';
  my $r = '';

  foreach $hit (@{$blast_parse->{ hits}}) {
40    $overlap = min( $end, $hit->{ qend}) - max( $start, $hit->{
    qstart});
    if( !$hit->{ self} && $overlap > 0 ) {
      $qoffset = max( 0, $start - $hit->{ qstart}) + 1;
      $l = ' ' x max( ($hit->{ qstart} - $start), 0);
45      $r = ' ' x max( ($end - $hit->{ qend}), 0);
      $$qseq = $l . substr( $hit->{ queryseq}, $qoffset, $overlap)
      . $r;
      $mseq = $l . substr( $hit->{ matchseq}, $qoffset, $overlap)
      . $r;
50      $tseq = $l . substr( $hit->{ targetseq}, $qoffset, $overlap)
      . $r;
      push @hits, ( "overlap", $overlap,
        "qseq", $qseq, "mseq", $mseq, "tseq", $tseq,
        "tname", $hit->{ tname} );
55    }
  }
  @hits;
}

```

```

sub match_score($$$$){
  my ($param, $blast_parse, $seq_start, $seq_end) = @_;
  my $myhit;
5  my $nm;
  my $max_match = 0;
  my $max_stretch = 0;
  my $l;
  my $max_stretch_score;
10  my $max_match_score;

  my @hits = find_oligo_hits( $blast_parse, $seq_start,
    $seq_end);

15  foreach $myhit (sort {$b->{ overlap} <=> $a->{ overlap}}
    @hits) {
    $_ = $myhit->{ mseq};
    while( /(\\|+)/g ) {
      $l = length($l);
20      if( $l > $max_stretch) {
        $max_stretch = $l;
      }
    }
    $nm = s/\\|/./g;
25    if( $nm > $max_match) { $max_match = $nm; }
    $max_match_score = squash(
      $param->{ oligod_param}{ max_match}{ squash_factor} *
      ( $param->{ oligod_param}{ max_match}{ cutoff} -
30    $max_match) );
    $max_stretch_score = squash(
      $param->{ oligod_param}{ max_stretch}{ squash_factor} *
      ( $param->{ oligod_param}{ max_stretch}{ cutoff} -
35    $max_stretch) );
    #print "MATCH $max_stretch, $max_match\n";

    ( ( "score", $max_match_score, "raw", $max_match),
      ( "score", $max_stretch_score, "raw", $max_stretch));
40  }

sub find_oligo($$$$$$){
  my ($glb, $seq, $position_score, $blast_parse, $param,
    $seq_number) = @_;
45  my $i;
  my $l = length($seq->{seq}) - $param->{ oligo_length};
  my $s;
  my $temp = 0;
  my $temp_tm = 0;
50  my $tm_dan = ( "temp", 0);
  my $tm;
  my $self_score = 0;
  my $db_score;
  my @result;
55  my $min_hit_score_score = 999999999;
  my @oligo_center = ();
  my $c = 0;
  my $dist;

```

```

my $this_oligo_overlaps= 0;
my @hits;
my $hitcount      = 0;
my $nm;
5 my $myhit;
my $lnaseq;
my $evidence;
my $count;
10 my $max_match;
my $seq_start;
my $seq_end;
my $oligo_property;
my @seqs;
my @lnaseqs;
15 my @dummy;
my @self_hyp;
my $self_hyp_batch = 1;
my $tm_batch      = 1;
my $tm_result;
20 my $target_mask;

if( $param->{ oligod_param}{ target_struct}{ weight} ) {
25   $target_mask = target_struct( $glb, $param, $seq);
}

for( $seq_start=1; $seq_start<$l; $seq_start++) {
   $seq_end      = $seq_start + $param->{ oligo_length};
   $s            = substr( $seq->{ seq}, $seq_start,
30     $param->{ oligo_length});
   $i            = $seq_start - 1;
   $seqs[ $i]    = $s;
   $lnaseqs[ $i] = capitalize( $s, $param->{ cphase},
35     $param->{ cfreq}, $param->{ end_spike_len});
}

if( $self_hyp_batch) {
   if( $param->{ oligod_param}{ self_hyp}{ weight} ) {
      if( $param->{ lna}) {
40         @self_hyp      = self_hyp( $glb, $param, @lnaseqs);
      }
      else {
         @self_hyp      = self_hyp( $glb, $param, @seqs);
45     }
   }
}

if( $tm_batch) {
   if( $param->{ oligod_param}{ tm_min}{ weight} &&
50     $param->{ oligod_param}{ tm_max}{ weight} ) {
      $tm      = tm( $glb, $param, @lnaseqs);
   }
}

55 for( $seq_start=1; $seq_start<$l; $seq_start++) {
   $i      = $seq_start - 1;
   $lnaseq = $lnaseqs[ $i];
   $seq_end = $seq_start + $param->{ oligo_length};
}

```

```

$oligo_property = ( "seq", $seqs[ $i], "start", $seq_start,
                    "end", $seq_end);
$ss = $seqs[ $i];
if( $param->( oligod_param){ hit_score}{ weight} ) {
5   $oligo_property->( hit_score) =
    hit_score( $position_score, $seq_start, $seq_end,
$param);
}

10   if( $param->( oligod_param){ target_struct){ weight} ) {
    $oligo_property->( target_struct) =
    target_struct_score( $target_mask, $seq_start, $seq_end,
$param);
}

15   if( $param->( oligod_param){ max_match){ weight} ||
    $param->( oligod_param){ max_stretch){ weight} ) {
    ($oligo_property->( max_match), $oligo_property->(
20   max_stretch))
    =
    match_score( $param, $blast_parse, $seq_start, $seq_end);
}

    if( $param->( oligod_param){ self_match){ weight} ) {
25     if( $param->( lna)) {
        $oligo_property->( self_match)= self_match( $glb, $param,
$lnaseq);
    }
    else {
30     $oligo_property->( self_match)= self_match( $glb, $param,
$s);
    }
}

35   if( $param->( oligod_param){ self_hyp){ weight} ) {
    if( $self_hyp_batch ) {
        $i = $seq_start - 1;
        $oligo_property->( self_hyp) = $self_hyp[ $i];
    }
40   else {
        if( $param->( lna)) {
            ($oligo_property->( self_hyp), @dummy) = self_hyp(
$glb, $param, ($lnaseq));
        }
45     else {
        ($oligo_property->( self_hyp), @dummy) = self_hyp(
$glb, $param, ($s));
    }
}

50   }

    if( $param->( oligod_param){ tm_dan_min){ weight} &&
        $param->( oligod_param){ tm_dan_max){ weight} ) {
55     ($oligo_property->( tm_dan),
        $oligo_property->( tm_dan_min), $oligo_property->(
tm_dan_max)) =
        tm_dan( $glb, $param, $s);

```

```

    )

    if( $param->( oligod_param){ tm_min}{ weight} &&
        $param->( oligod_param){ tm_max}{ weight} ) {
5      if( $tm_batch ) {
        $i = $seq_start - 1;
        $tm_result      = $tm->{$i};
        $oligo_property->( tm)      = $tm_result->( tm);
        $oligo_property->( tm_min) = $tm_result->( tm_min);
10      $oligo_property->( tm_max) = $tm_result->( tm_max);
      }
      else {
        $tm_result      = tm( $glb, $param, ($lnaseq));
        $oligo_property->( tm)      = $tm_result->[0]{ tm};
15      $oligo_property->( tm_min) = $tm_result->[0]{ tm_min};
        $oligo_property->( tm_max) = $tm_result->[0]{ tm_max};
      }
    }

20    if( $param->( oligod_param){ ggg){ weight} ) {
        $oligo_property->( ggg) = ggg_score( $param, $s);
    }

25    if( $param->( oligod_param){ cg){ weight} ) {
        $oligo_property->( cg) = cg_score( $s);
    }

    $oligo_property->( score) =
30      oligo_property_score( $param, $oligo_property);

    push @result, $oligo_property;
  )

35  $i = 0;
  foreach $oligo_property ( sort {$b->( score) <=> $a->( score)}
    @result) {
    $this_oligo_overlaps = 0;
    if( $param->( mindist) ) {
40      $c = $oligo_property->( start) +
        ($oligo_property->( end) - $oligo_property->( start)) / 2;
        foreach (@oligo_center) {
          $dist = abs( $_ - $c);
          if( $dist < $param->( mindist) ) {
45            $this_oligo_overlaps = 1;
            last;
          }
        }
    }
  }

50  #print "Oligo center is $c, overlap is $this_oligo_overlaps\n";
  if( ! $this_oligo_overlaps) {
    $count++;
    if( $glb->( view) =~ /PROBE/ ) {
      print "<PROBE>\n";
55      print "seqno: $seq_number oligono: $count\n";
    }
  }
  # $temp = melt_temp( $oligo_property->( seq));

```

```

    $lnaseq      = capitalize( $oligo_property->( seq), $param-
>(cphase),
                                $param->( cfreq), $param->(
end_spike_len));
5    # $temp_tm = tm_pred( $lnaseq);
    # $tm_dan    = tm_dan( $glb, $param, $oligo_property->( seq));
    # $self_score = self_match( $glb, $param, $oligo_property-
>( seq));
    @hits = find_oligo_hits( $blast_parse, $oligo_property->(
10    start),
        $oligo_property->( end));

    $max_match = 0;
    foreach $myhit (sort ($b->( overlap) <=> $a->( overlap))
15    @hits) {
        $_ = $myhit->( mseq);
        $nm = s/|/./g;
        if( $nm > $max_match) { $max_match = $nm; }
    }
20    if( $glb->( view) == /SCORE_EVIDENCE/ ) {
        print_oligo_property( $param, $oligo_property);
    }

25    if( $glb->( view) == /PROBE/ ) {
        XXX printf( "score=%4.2f tm_ex=$temp_tm", $oligo_property-
>( score));
        printf( "score=%4.2f ", $oligo_property->( score));
        if( $param->( oligod_param){ tm_min}{ weight} &&
30        $param->( oligod_param){ tm_max}{ weight}) {
            #print " ", tm=$oligo_property->( tm_min){ raw}\n\n";
            print " ", tm=$oligo_property->( tm_min){ raw}";
        }
        else {
35        print "\n";
        }
        if( $param->( oligod_param){ tm_dan_min}{ weight} &&
        $param->( oligod_param){ tm_dan_max}{ weight}) {
            #print " ", tm_dan=$oligo_property->( tm_dan_min){ raw}\n\n";
40        print " ", tm_dan=$oligo_property->( tm_dan_min){ raw}\n\n";
        }
        else {
            print "\n\n";
        }
45        print "$lnaseq $oligo_property->( start) $oligo_property->(
end)\n";
        #)

    if( $glb->( view) == /HITS/ ) {
50        print "<HITS>\n";
        #        "$oligo_property->( dbmatch), ".
        #        "$window->( self_score), $max_match\n";
        $hitcount = 0;
        foreach $myhit (sort ($b->( overlap) <=> $a->( overlap))
55    @hits) {
            if( $hitcount >= $param->( maxhits)) {
                my $n = $#hits + 1;
                print "Only showing $param->( maxhits) hits of $n hits\n";
            }
        }
    }

```

```

        last;
    }
    #print "\n$myhit->( qseq)\n$myhit->( mseq)\n$myhit->(
tseq)\n";
5    $_ = $myhit->( mseq);
    $nm = s/\|/./g;
    printf( "$myhit->( mseq) %2d matches\n$myhit->( tseq)
$myhit->( tname)\n",
                                                $nm);
10    $hitcount++;
    }
    print "</HITS>\n";
    }
    if( $glb->( view) == /SELF_EVIDENCE/ ) {
15    print "<SELF_EVIDENCE>\n";
        if( $param->( oligod_param){ self_hyp}{ weight} ) {
            print "$oligo_property->( self_hyp){ evidence}";
        }
        if( $param->( oligod_param){ self_match){ weight} ) {
20    print "$oligo_property->( self_match){ evidence}";
        }
        print "</SELF_EVIDENCE>\n";
    }
    if( $glb->( view) == /TARGET_STRUCT_HIT/ ) {
25    print "<TARGET_STRUCT_HIT>\n";
        if( $param->( oligod_param){ target_struct){ weight} ) {
            print "$oligo_property->( target_struct){ evidence}";
        }
        print "\n</TARGET_STRUCT_HIT>\n";
30    }
    if( $glb->( view) == /PROBE/ ) {
        print "</PROBE>\n";
    }
    push @oligo_center, $c;
35    $i++;
    }
    if( $i > $param->( max_noligo) ) { last; }
    }
40    }

##### MAIN #####

45    sub main($){
        my ($glb) = @_;
        my $lastline;
        my $seq;
        my $count = 0;
50    my $position_score;
        my $blast_parse;
        my $hit;
        my $check_blast = 1;
        my $param;
55    my $starttime;
        my $date;
        my $runtime;
        my $opt;

```



```

my $popt;
local *INFILE;
my $oligo_design = 0;

5
if(0) {
    $seq = read_file( "out.blast");
    #print $seq;
    parse_blast( $seq);
10    exit( 0);
}

($opt, $popt) = parse_argv( $glb);

15    read_conf( $glb);

    $param = initialize_param( $popt);

20    mkdir( $glb->( tmpdir));

25    foreach (@ARGV) {
        if( $_ eq "-") {
            *INFILE = *STDIN;
        }
        else {
30            open( INFILE, "<$_") || die "Could not open $_ $!";
        }

        while( $seq = read_fasta( \ $lastline, \*INFILE) ) {
35            $count++;
            if( $opt->( capitalize)) {
                $seq->( name) = $seq->( name) . "_lna";
                $seq->( seq) = capitalize( $seq->( seq),
                    $param->( cphase), $param->( cfreq),
40                $param->( end_spike_len));
                print_fasta( $seq, \*STDOUT);
            }
            elsif( $opt->( rev)) {
45                $seq->( seq) = rev( $seq->( seq));
                $seq->( name) = $seq->( name) . "_rev";

                $seq->( seq) = capitalize( $seq->( seq), 2, 3, 0);

                print_fasta( $seq, \*STDOUT);
50            }
            elsif( $opt->( comp)) {
                $seq->( seq) = comp( $seq->( seq));
                $seq->( name) = $seq->( name) . "_comp";
                print_fasta( $seq, \*STDOUT);
55            }
            elsif( $opt->( revcomp)) {
                $seq->( seq) = revcomp( $seq->( seq));
                $seq->( name) = $seq->( name) . "_revcomp";
            }
        }
    }

```

```

    print_fasta( $seq, \*STDOUT);
}
elseif( $opt->( self_score)) {
    my ($evidence, $self_score) =
5       self_hyp( $glb, $param, $seq->( seq));
    print( "score = $self_score\n");
}
else{
    $starttime = time();
10    if( ! $oligo_design) {
        my ($sec,$min,$hour,$mday,$mon,$year) = (localtime(
            $starttime))[0..5];
        $year += 1900;
        $mon++;
15        $date = "$year $mon $mday $hour:$min:$sec";
        $oligo_design = 1;
        print "<PROBE_RUN $date version=$glb->( version)>\n";
        if( $glb->( view) =~ /PARAMETER/ ) {
            print dump_param( $glb, $param);
20        }
    }
    if( $param->( oligo_sense) eq "reverse") {
        $seq->( seq) = revcomp( $seq->( seq));
    }
25    print "<PROBE_DESIGN>\n";
    if( $check_blast && $param->( fastadb)) {
        my @fastadbs = split( " ", $param->( fastadb));
        my $bdb;
        foreach (@fastadbs) {
30            $bdb = formatblastdb( $glb, $_, "F");
            $param->( blastdb) .= " " . $bdb;
        }
        $check_blast = 0;
    }
35    ($position_score, $blast_parse) = position_score( $glb,
        $param, $seq);

    if( $glb->( view) =~ /FASTA/ ) {
40        print "<FASTA>\n";
        print_fasta( $seq, \*STDOUT);
        print "</FASTA>\n";
    }

    find_oligo( $glb, $seq, $position_score, $blast_parse,
45    $param, $count);

    $runtime = time() - $starttime;
    if( $glb->( view) =~ /RUNTIME/ ) {
50        print "<RUNTIME>$runtime s</RUNTIME>\n";
    }
    print "</PROBE_DESIGN>\n";
}
close INFILE;
55 }
if( $oligo_design) {
    print "</PROBE_RUN>\n";
}

```

```
    }  
5  main( $glb);
```

Figure 28B

19 MAJ 2003

Modtaget

```

/*
   $Id: dyp.c,v 1.15 2002/07/29 06:28:00   nt Exp $

5  DESCRIPTION
    Loading sequences:
    -f fastafile
    -seq sequence

10  Set min_score to 0 to see all structures
    -min_score min_score

    Size of sliding window default is INT_MAX
    -depth depth

15  Number of results to show
    -max_res max_res

    -secondary_structure
20  Calculates the secondary structure of an oligo

    -self_annealing
    Calculates the binding energy between an oligo and
    it self.

25  -hybridization
    Calculates the binding energy between two different
    oligos, or an oligo and its target.

30  AUTHOR
    Copyright Niels Tolstrup 2002
    Exiqon
    */

35  #include <stdio.h>      /* fprintf, rand...*/
    #include <stdlib.h>    /* calloc...*/
    #include <stdarg.h>    /* va_list... */
    #include <string.h>    /* strlen...*/
40  #include <limits.h>    /* USHRT_MAX... */
    #include <math.h>      /* pow */
    #include <ctype.h>     /* tolower */
    #include "getopt.h"    /* getopt_long */

45  /* export MALLOC_TRACE=dyp_memtrace
    mtrace dyp $MALLOC_TRACE
    #include <mcheck.h>    mtrace
    */

50  /***** DEFINES *****/

    #define THEVERSION "1.2 2002-10-11-15-46"
    #define TRACEBACK_INT unsigned short
55  #define TRACEBACK_INT_MAX USHRT_MAX

    #define HI printf( "Hello\n");

```

```

#define SECONDARY_STRUCTURE 1
#define SELF_ANEALING 2
#define HYBRIDIZATION 4
#define TM 8

5  /* Increase the stack by this amount in case of overflow */
#define STACK_CHUNK 255

/* Number of sequences to allocate space for before realloc */
10 #define SEQS_CHUNK 255

/* Size of sequence to allocate space for before realloc */
#define SEQ_CHUNK 1024

15 /* Longest possible line in a matrix that can be read */
#define MAX_LINE_LEN 5000

#define MIN(a,b) (a<b?a:b)
#define MAX(a,b) (a>b?a:b)

20

/***** GLOBAL VARIABLES *****/

25 int verbose = 0;
char* program_name;

/***** STRUCTURES *****/

30

typedef struct sequences_t {
    int nseq;
    int maxnseq;
35 char **seqs;
} sequences_type;

typedef struct dynamic_str_t {
40 int len;
    int maxlen;
    char *s;
} dynamic_str_type;

45

typedef struct pair_t {
    TRACEBACK_INT i;
    TRACEBACK_INT j;
50 } pair_type;

typedef struct stack_t {
    TRACEBACK_INT sp;
    TRACEBACK_INT size;
55 pair_type *stack;
} stack_type;

```

```

typedef struct tokenarray_t {
    int      ntok;
    int      maxntok;
    char      **tok;
5      } tokenarray_type;

typedef struct matrix_t {
10     char      *seq1;
    char      *seq2;
    int l1;
    int l2;
    int max_size;
15     int *d;
    } matrix_type;

20 typedef struct score_rec_t {
    int      alph_len;
    int      gap_start;
    int      gap_cont;
    int      loop_score;
25     int      match_cont_factor;
    int      match_threshold;
    int      min_strong_ident_score;
    int      min_ident_score;
    int      min_sim_score;
30     char      *alph;
    int      *mat;
    } score_rec_type;

35 typedef struct param_t {
    char      *alph;
    int      *mathyp;
    int      depth;
    int      min_score;
40     int      max_res;
    sequences_type *seqs;
    sequences_type *seqs2;
    score_rec_type *score_rec;
    } param_type;
45

/*****      SYSTEM      *****/

50 void die( char *fmt, ...){
    va_list ap;
    fprintf( stderr, "Uhoh: ");
    va_start( ap,      fmt);
    vfprintf( stderr, fmt, ap);
55     va_end( ap);
    fprintf( stderr, "\n");
    exit( 1);
}

```

```

void usage( char *fmt, ...){
    va_list ap;
5   va_start( ap,    fmt);
    vfprintf( stderr, fmt, ap);
    va_end( ap);
    fprintf( stderr,
10      "\nUsage: dyp -h,--help\n"
        "    -v,--verbose\n"
        "    -r,--version\n"
        "    -u,--secondary_structure\n"
        "    -a,--self_anealing\n"
        "    -y,--hybridization\n"
15      "    -t,--tm\n"
        "    -o,--output filename\n"
        "    -i,--input filename\n"
        "    -j,--input2 filename2\n"
        "    -s,--seq sequence\n"
20      "    -z,--seq2 sequence2\n"
        "    -d,--depth depth\n"
        "    -n,--min_score score\n"
        "    -p,--max_res    n\n"
        "    -m,--matrix matfn\n"
25      "    -l,--allolig length_of_oligo\n"
        "    -f,--spikefreq freq\n"
        "    -e,--sample n_samples\n");
    fprintf( stderr, "\n");
    exit( 1);
30 }

void hi( char *fmt, ...){
    va_list ap;
35   fprintf( stdout, "Hi there " );
    va_start( ap,    fmt);
    vfprintf( stdout, fmt, ap);
    va_end( ap);
    fprintf( stdout, "\n");
40 }

void *salloc( int nobj, int size){
    void *mem;
45   mem = calloc(  nobj, size);
    /*printf( "calloc: allocating  %d objects of size %d bytes\n",
        nobj, size);*/
    if( mem == NULL){
50      die( "Could    not allocate %d    x %d bytes\n", nobj,
size);
    }
    return mem;
}

55 FILE *s fopen( char *fn, const char *mode){
    FILE *file;
    file = fopen(    fn, mode);

```

```

    if( file == NULL )
        die( "Could not open %s for %s.", fn, mode);
    return file;
}

/***** STRING FUNCTIONS *****/

10 char *empty_str( int len){
    char *str;
    str = calloc( len+1, sizeof( char));
    memset( str, ' ', len);
    str[len] = 0;
15 return str;
}

char *reverse( char *str){
20 int i, len;
    char c;

    len = strlen( str);
    for( i=0; i<len/2; i++){
25 c = str[i];
        str[i] = str[(len-1) - i];
        str[(len-1) - i] = c;
    }
    return str;
30 }

char *lower( char *s) {
    int i, l = strlen( s);
35 for( i=0; i<l; i++)
        s[i] = tolower( s[i]);
    return s;
}

40 char tocomp( char c) {
    char fr_s[35] = "acgtumrwsykvhdbxnACGTUMRWSYKVHDBXN";
    char to_s[35] = "tgcaakywsrmbdhvxxTGCAAKYWSRMBDHVXX";
    char *p;
45 char cc;

    p = strchr( fr_s, c);
    if( p ) {
        p = to_s + (p - fr_s);
50 cc = *p;
    }
    else
        cc = 'x';
    return cc;
55 }

char *comp( char *s) {

```



```

    int i, l = strlen( s );
    for( i=0; i<l; i++)
        s[i] = tocomp( s[i] );
    return s;
5  }

    /***** I/O of strings *****/

10 void *print_lines( FILE *outfile, char **str, int nstr ){
    int linelen = 80;
    int i, j, p, len;

15     if( nstr > 0 ) {
        len = strlen( str[0] );

        for( j=0; j<nstr; j++) {
            if( strlen( str[j] ) != len ) {
20                 die( "String len = %d differs from first stringlen %d
                    %s\n",
                        strlen( str[j] ), len, str[j] );
            }
        }
        p = 0;

        while( p < len ) {
            for( j=0; j<nstr; j++) {
                for( i = p; i < p + linelen && i < len; i++) {
30                     putc( str[j][i], outfile );
                }
                putc( '\n', outfile );
            }
            p += linelen;
35             putc( '\n', outfile );
        }
    }
}

40 /***** STACK *****/

    stack_type *init_stack( TRACEBACK_INT size ) {
45     stack_type *stack;

    stack = calloc( 1, sizeof( stack_type ) );
    stack->sp = 0;
    stack->size = size;
50     stack->stack = calloc( size, sizeof( pair_type ) );

    return stack;
}

55 void *free_stack( stack_type *stack ){
    free( stack->stack );
    free( stack );
}

```

```

    }

    void print_pair( pair_type pair){
5      printf( "%4d,    %4d\n",  pair.i,  pair.j);
    }

    void push( stack_type *stack, pair_type  pair) {
10      int  new_size;
        if( stack->sp < stack->size ) {
            stack->stack[ stack->sp++] = pair;
            if(0){printf("pushed :"); print_pair( pair);}
        }
15      else {
            new_size = (stack->size + STACK_CHUNK) * sizeof( pair_type);
            if( stack->stack = realloc( stack->stack, new_size)) {
                stack->size
                    = stack->size + STACK_CHUNK;
                stack->stack[ stack->sp++] = pair;
20            }
            else {
                die( "Failed to realloc stack to %d    bytes",
                    new_size);
            }
25        }
    }

    pair_type setpair( int i, int j) {
30      pair_type pair;
        pair.i = i;
        pair.j = j;
        return pair;
    }
35

    pair_type pop( stack_type *stack) {
        if( stack->sp > 0 )
            return stack->stack[    --(stack->sp)];
40      else
        return setpair( TRACEBACK_INT_MAX, TRACEBACK_INT_MAX);
    }

    void test_stack(){
45      stack_type *stack;
        int    i;

        stack = init_stack( STACK_CHUNK);
50
        for( i=0; i<30; i++)
            push( stack,    setpair( i, i+1));

        for( i=0; i<30; i++)
55      print_pair( pop( stack));

        free_stack( stack);
    }

```

```

/***** DYNAMIC STRINGS *****/

5  free_dynamic_str_type( dynamic_str_type *seq) {
    free( seq->s);
    free( seq);
10 }

    dynamic_str_type *new_dynamic_str_type(){
        dynamic_str_type *seq;

15     seq = calloc( 1, sizeof( dynamic_str_type));
        seq->len = 0;
        seq->maxlen = SEQ_CHUNK;
        seq->s = calloc( seq->maxlen, sizeof( char));
        seq->s[0] = '\0';
20     return seq;
    }

    dynamic_str_type *addchar( dynamic_str_type *seq, int c) {
25     if( seq == NULL) {
        seq = new_dynamic_str_type();
    }
    if( seq->len + 1 >= seq->maxlen ) {
        seq->maxlen += SEQ_CHUNK;
30     if( ! (seq->s = realloc( seq->s, seq->maxlen * sizeof(
char))) )
        die( "Failed to realloc seq to %d chars", seq->maxlen);
    }
    seq->s[ seq->len++] = c;
35     return seq;
    }

/***** OLIGO FUNCTIONS *****/

40
    long n_nmer( int n_mer, char *alph){
        int alphlen = strlen( alph);
        long n_nmer = 0;

45     if( alphlen == 0 )
        n_nmer = 0;
    else
        n_nmer = pow( alphlen, n_mer);

50     return n_nmer;
    }

55 char *make_nmer( long number, int n_mer, char *alph) {
    long i, j, n = number;
    int alphlen = strlen( alph);
    char *seq;

```

```

    seq = empty_str( n_mer);

    if( alphlen == 0 )
5      strcpy( seq, "-");
    else {
        for( j=0; j<n_mer; j++){
            i = n % alphlen;
            n = n / alphlen;
10      seq[j] = alph[i];
        }
        seq = reverse( seq);
15      return seq;
    }

char *make_random_nmer( int n_mer, int spikefreq, char *alph) {
20      long i, j;
      int alphlen = strlen( alph);
      int halflen = alphlen / 2;
      char *seq;
      char phase = 0;
25      seq = empty_str( n_mer);

      if( spikefreq)
          phase = (rand() * spikefreq)/(RAND_MAX+1);
30

      if( alphlen == 0 )
          strcpy( seq, "-");
      else {
35          if( spikefreq)
              for( j=0; j<n_mer; j++){
                  i = (rand() * halflen)/(RAND_MAX+1);
                  if( (j+phase) % spikefreq == 0)
                      i += halflen;
40          seq[j] = alph[i];
              }
          else
              for( j=0; j<n_mer; j++){
                  i = (rand() * alphlen)/(RAND_MAX+1);
45          seq[j] = alph[i];
              }
          }
      return seq;
50  }

/***** MATRIX I/O *****/

55  score_rec_type *init_score_rec_type( int gap_start, int
      gap_cont,
      int loop_score,    int match_threshold,

```

```

        int match_cont_factor,
        int min_strong_ident_score,
        int min_ident_score, int min_sim_score,
        char *alph, int *mat){
5   score_rec_type *score_rec;

    score_rec = (score_rec_type *) calloc( 1, sizeof(
score_rec_type));
    score_rec->alph_len = strlen( alph);
10   score_rec->gap_start = gap_start;
    score_rec->gap_cont = gap_cont;
    score_rec->loop_score = loop_score;
    score_rec->match_threshold = match_threshold;
    score_rec->match_cont_factor = match_cont_factor;
15   score_rec->alph = alph;
    score_rec->mat = mat;
    score_rec->min_strong_ident_score = min_strong_ident_score;
    score_rec->min_ident_score = min_ident_score;
    score_rec->min_sim_score = min_sim_score;
20   return score_rec;
}

25 void free_score_rec_type( score_rec_type *score_rec){
    free( score_rec->alph);
    free( score_rec->mat);
    free( score_rec);
30 }

void free_param_type( param_type *param){
    free_score_rec_type( param->score_rec);
    free( param->seqs->seqs);
35   free( param->seqs);
    free( param->seqs2->seqs);
    free( param->seqs2);
    free( param);
40 }

void validate_score_rec_type( score_rec_type *score_rec, int
min, int max){
    int i, j;
45   if( strlen( score_rec->alph) < 1) { die( "No alphabet\n"); }
    if( strlen( score_rec->alph) != score_rec->alph_len) {
        die( "alph length mismatch %d,%d\n", strlen( score_rec-
>alph),
50         score_rec->alph_len);
    }
    for( i=0;i<score_rec->alph_len;i++){
        for( j=0;j<score_rec->alph_len;j++){
            if( score_rec->mat[i*score_rec->alph_len+j]<min )
                die( "Matrix out of range min=%d", min);
55         if( score_rec->mat[i*score_rec->alph_len+j]>max )
            die( "Matrix out of range %d max=%d",
                score_rec->mat[i*score_rec->alph_len+j], max);
        }
    }
}

```

```

    )
    if( score_rec->gap_start > max ) die( "gap_start out of range
max");
    if( score_rec->gap_start < min ) die( "gap_start out of range
5 max");
    if( score_rec->gap_cont > max ) die( "gap_cont out of range
max");
    if( score_rec->gap_cont < min ) die( "gap_cont out of range
max");
10    if( score_rec->loop_score > max ) die( "loop_score out of
range max");
    if( score_rec->loop_score < min ) die( "loop_score out of
range max");
    )
15

void print_mat(    char *seq1, char *seq2, int *mat){
    int    i, j;
    int    len1 = strlen( seq1);
20    int    len2 = strlen( seq2);
    int    p_num = 1;

    if( p_num == 1) {
        printf( " ");
25        for( i=0;i<len1;i++){
            printf( " %2d ", i);
        }
        printf( "\n ");
    }
30    printf( " ");
    for( i=0;i<len1;i++){
        printf( " %c ", seq1[i]);
    }
    printf( "\n\n");
35    for( j=0;j<len2;j++){
        if( p_num == 1) {
            printf( "%2d %c ", j, seq2[j]);
        }
        else {
40            printf( "%c ", seq2[j]);
        }
        for( i=0;i<len1;i++){
            printf( "%5d ", mat[j*len1+i]);
        }
45        printf( "\n");
    }
}

50 /***** TOKENIZE A LINE *****/

void addtoken( char* token, tokenarray_type *tokena){
    tokena->ntok++;
55    if( tokena->ntok > tokena->maxntok ) {
        tokena->maxntok += 100;
        if( ! (tokena->tok =

```

```

        realloc( tokena->tok, tokena->maxntok * sizeof(
            char))) )
        die( "Failed to realloc tokenarray to %d tokens",
tokena->maxntok);
5    }
    tokena->tok[ tokena->ntok - 1] = token;
}

10 tokenarray_type *alloc_tokenarray_type(){
    tokenarray_type *tokena;

    tokena = calloc( 1, sizeof( tokenarray_type));
    tokena->maxntok = 30;
15    tokena->ntok = 0;
    tokena->tok = calloc( tokena->maxntok, sizeof( char*));
    return tokena;
}

20 void *free_tokenarray_type( tokenarray_type *tokena){
    free( tokena->tok);
    free( tokena);
}

25 void *tokenize( char *s, tokenarray_type *tokena){
    char *token;

30    tokena->ntok= 0;
    token = strtok( s, " \n");
    if( token )
        addtoken( token, tokena);
    while( token = strtok( NULL, " \n"))
35    addtoken( token, tokena);
}

int tokenarray_is_num_list( tokenarray_type *tokena){
40    int i = 0;
    char *endp;
    int res = 0;

    while( i<tokena->ntok && strtol( tokena->tok[i], &endp, 10) ==
45    i) {
        if( *endp != 0 )
            res = 0;
        i++;
    }
50    if( i == tokena->ntok)
        res = i;
    else
        res = 0;
    return res;
55 }

char *tokenarray2str( tokenarray_type *tokena){

```

```

    int i=0;
    char *alph;

    alph = calloc( tokena->ntok + 1, sizeof( char));
5   for( i=0; i<tokena->ntok; i++) {
        if( strlen( tokena->tok[i]) != 1)
            die( "Sorry, the word %s was found were a char was
expected.",
                                tokena->tok[i]);
10      alph[i] = *tokena->tok[i];
        }
        alph[i+1] = 0; /* EOL */

        return alph;
15  }

int tokenarray2dat( tokenarray_type *tokena, int with_num,
                    matrix_type *mt, char *c){
20  int i = 0;
    int j = 0;
    char *endp;
    int err = 0;
    int *dat;
25
    if( with_num ) {
        if( atoi( tokena->tok[i]) != mt->l2)
            err = 1;
            i++;
30    }
    if( strlen(tokena->tok[i]) != 1)
        err = 2;
        *c = *tokena->tok[i];
        i++;
35
    if( tokena->ntok - i != mt->l1 )
        err = 3;
    else {
        dat = calloc( mt->l1, sizeof( int));
40    for( ; i<tokena->ntok; i++) {
        dat[j] = strtol( tokena->tok[i], &endp, 10);
        if( *endp != 0 )
            err = j + 10;
            j++;
45    }
    }
    if( !err) {
        if( mt->l1 * mt->l2 > mt->max_size ) {
            mt->max_size += 100;
50        if( ! (mt->d =
            realloc( mt->d, mt->max_size * sizeof( int))) )
            die( "Failed to realloc matrix to %d int's", mt-
            >max_size);
        }
65    for( j=0; j<mt->l1; j++)
        mt->d[mt->l1 * mt->l2 + j] = dat[j];
        mt->l2++;
    }
}

```



```

    free( dat);
    return err;
}

5  /***** I/O score_rec_type *****/

matrix_type *read_mat( FILE *file){
10  char    *s, *res;
    int     n = MAX_LINE_LEN;
    int     col_name_found = 0;
    int     numbers_found = 0;
    tokenarray_type *tokena;
15  dynamic_str_type *seq = NULL;
    char    c;
    matrix_type *mt;

    mt = calloc( 1, sizeof( matrix_type));
20  mt->max_size = 100;
    mt->d = calloc( mt->max_size, sizeof( int));
    mt->l2 = 0;

    tokena = alloc_tokenarray_type();

25  s = calloc( n + 1, sizeof( char));
    seq = new_dynamic_str_type();

    do {
30  res = fgets( s, n + 1, file);
        if( res != NULL ) {
            tokenize( s, tokena);
            /*for(i=0;i<tokena->ntok;i++) printf( "%d %s\n", i, tokena-
>tok[i]);*/
35  if( !col_name_found ) {
                /* Is it a column number line? */
                if( !numbers_found && (mt->l1 = tokenarray_is_num_list(
tokena))) {
40  numbers_found = 1;
                }
                else {
                    /* Is it a column name line? */
                    mt->seq1 = tokenarray2str( tokena);
                    if( numbers_found )
45  if( mt->l1 != strlen( mt->seq1))
                        die( "Expected a string of length %d but found %s",
mt->l1,
                                mt->seq1);
                    else
50  mt->l1 = strlen( mt->seq1);
                    col_name_found = 1;
                }
            }
            else {
55  /* skip empty lines */
                if( tokena->ntok > 0 ) {
                    /* Is it a data line? */

```

```

        if( tokenarray2dat( tokena, numbers_found, mt, &c) == 0)
        {
            seq = addchar(    seq, c);
        }
5       else {
            res = NULL;
        }
    }
10    }
    ) while( res);

    mt->seq2 = calloc( strlen( seq->s) + 1, sizeof( char));
    strcpy( mt->seq2, seq->s);
15    free_dynamic_str_type( seq);

    free( . s);
    free_tokenarray_type( tokena);
    return mt;
20 }

void print_score_rec_type( score_rec_type *score_rec){
    printf( "alph_len   %d\n", score_rec->alph_len);
25    printf( "gap_start %d\n", score_rec->gap_start);
    printf( "gap_cont   %d\n", score_rec->gap_cont);
    printf( "alph       %s\n", score_rec->alph);
    print_mat( score_rec->alph, score_rec->alph, score_rec->mat);
30 }

void read_score_rec_type( char *fn, score_rec_type *score_rec){
    FILE *file;
    matrix_type *mt;
35
    file = fopen( fn, "r");
    fscanf( file, "alph_len   %d\n", &(score_rec->alph_len));
    fscanf( file, "gap_start  %d\n", &(score_rec->gap_start));
    fscanf( file, "gap_cont   %d\n", &(score_rec->gap_cont));
40    fscanf( file, "min_strong_ident_score %d\n", &(score_rec->min_strong_ident_score));
    fscanf( file, "min_ident_score %d\n", &(score_rec->min_ident_score));
    fscanf( file, "min_sim_score %d\n", &(score_rec->min_sim_score));
45    fscanf( file, "alph       %s\n", score_rec->alph);
    mt = read_mat( file);
    if( mt->l1 != mt->l2)
        die("Matrix not quadratic %d %d", mt->l1, mt->l2);
50    if( strcmp( mt->seq1, mt->seq2) != 0)
        die("Matrix sequences not identical %s %s", mt->seq1,
mt->seq2);
    if( strcmp( mt->seq1, score_rec->alph) != 0)
        die("Matrix alphabet differs from alph %s %s", mt->seq1,
55 score_rec->alph);
    if( score_rec->mat )
        free( score_rec->mat);
    score_rec->mat = mt->d;

```

```

    free( mt->seq1);
    free( mt->seq2);
    free( mt);
    fclose( file);
5  )

/***** MATRIX *****/

10 int *seq2idx_seq( char *seq, score_rec_type *score_rec) {
    char    c[2];
    int     i, l;
    int     *idx_seq;
15  int     id;

    l = strlen( seq);
    idx_seq = calloc( l, sizeof( int));

20  for( i=0; i<l; i++) {
        c[0] = seq[i];
        c[1] = 0;
        id = strcspn( score_rec->alph, c);
        if( id >= score_rec->alph_len) {
25      fprintf( stderr,
            "Sorry, the character %c was not found in the alphabet
            %s\n",
                                seq[i], score_rec->alph);
        id = score_rec->alph_len - 1;
30      }
        idx_seq[i] = id;
    }
    return idx_seq;
}

35

int matidx( int i, int j, int l1, int l2){
    if( i<0 ) die( "matrix index i (%d) less than zero", i);
    if( i>=l1 ) die( "matrix index i (%d) too large (%d)", i,
40  l1);
    if( j<0 ) die( "matrix index j (%d) less than zero", j);
    if( j>=l2 ) die( "matrix index j (%d) too large (%d)", j,
    l2);
    return j*l1 + i;
45  }

int matsize( int l1, int l2){
50  return l1*l2;
}

int m( int i, int j, matrix_type *mat) {
    return mat->d[ matidx( i, j, mat->l1, mat->l2)];
55  }

void print_annot( char *seq, char *annot){

```

```

    printf("%s\n", seq);
    printf("%s\n", annot);
}

5 char match_sym( char a, char b, score_rec_type *score_rec)
{
    int i1, i2;
    char *s;
10 int score;

    s = strchr( score_rec->alph, a);
    if( s == NULL) return ' ';
    i1 = s - score_rec->alph;
15 s = strchr( score_rec->alph, b);
    if( s == NULL) return ' ';
    i2 = s - score_rec->alph;

    score = score_rec->mat[ i2*score_rec->alph_len+i1];

20 if( score >= score_rec->min_strong_ident_score )
    return '|';
    else if( score >= score_rec->min_ident_score )
    return ':';
25 else if( score >= score_rec->min_sim_score )
    return '.';

    return ' ';
}
30

/***** ALIGN - self association *****/

35 void align_global_traceback( FILE *outfile, param_type * param,
    char *seq1, char *seq2, int *path){
    int l1 = strlen( seq1);
    int l2 = strlen( seq2);
    int done = 0;
40 int is = 0;
    int i, j;
    int i1, i2;
    char *qseq, *mseq, *tseq;

45 if( l1>0 && l2>0 ) {

    qseq = calloc( l1+l2+1, sizeof( char));
    mseq = calloc( l1+l2+1, sizeof( char));
    tseq = calloc( l1+l2+1, sizeof( char));

50 i = l1 - 1;
    j = l2 - 1;
    i1 = l1 - 1;
    i2 = l2 - 1;

55 do {
    if( verbose )
    printf( "%d %d\n", i, j);

```

```

        if( path[j*11+i] == 0 ) {
            if( i1 >= 0 ) {
2         qseq[is] = seq1[i1];
3         i1--;
4         }
5         else {
6         qseq[is] = '-';
7         mseq[is] = ' ';
8         }
9         if( i2 >= 0 ) {
10        tseq[is] = seq2[i2];
11        i2--;
12        }
13        else {
14        tseq[is] = '-';
15        mseq[is] = ' ';
16        }
17        /*
18        if( tseq[is] == qseq[is] ) { mseq[is] = ':'; }
19        else { mseq[is] = ' '; }
20        */
21        mseq[is] = match_sym( tseq[is], qseq[is], param-
22        >score_rec);
23        is++;
24        i--;
25        j--;
26        if( i<0 && j<0 ) {
27        done = 1;
28        }
29        }
30        else if( path[j*11+i] == 2 ) {
31        qseq[is] = '-';
32        mseq[is] = ' ';
33        tseq[is] = seq2[i2];
34        i2--;
35        is++;
36        j--;
37        }
38        else if( path[j*11+i] == 1 ) {
39        qseq[is] = seq1[i1];
40        i1--;
41        mseq[is] = ' ';
42        tseq[is] = '-';
43        is++;
44        i--;
45        }
46        if( i < 0 ) i = 0;
47        if( j < 0 ) j = 0;
48        if( verbose )
49        printf( "%d %d %c %c %c\n", i, j, qseq[is-1], mseq[is-1],
50        tseq[is-1]);
51        } while( !done);
52        qseq[is] = 0;
53        mseq[is] = 0;
54        tseq[is] = 0;
55
56        fprintf( outfile, "%s\n", reverse( qseq));

```

```

    fprintf( outfile, "%s\n", reverse( mseq));
    fprintf( outfile, "%s\n", reverse( tseq));

    free( qseq);
5    free( mseq);
    free( tseq);
}
)

10 int align_local_traceback( FILE *outfile, param_type      *param,

    int l1, int l2, int *idx_seq1, int *idx_seq2,
    char *mask_seq1, char *mask_seq2, char *seq1, char *seq2,
15    int *mat, int *path){

    stack_type      *stack;
    int              i, j, k;
    pair_type        pair;
20    char            *annot;
    matrix_type      *mt;
    int              score, maxscore, firstscore=0, thescore=INT_MAX,
    max_i, max_j;
    int              n_hyp      = 0;
25    char            *outstr[3];
    char            *qseq, *mseq, *tseq;

    stack            = init_stack( STACK_CHUNK);

30    mt              = calloc( 1, sizeof( matrix_type));
    mt->d             = mat;
    mt->l1             = l1;
    mt->l2             = l2;

35    qseq = calloc( l1+l2+1, sizeof( char));
    mseq = calloc( l1+l2+1, sizeof( char));
    tseq = calloc( l1+l2+1, sizeof( char));

    while( thescore >= param->min_score && n_hyp < param->max_res)
40    {
        maxscore = INT_MIN;
        for( j=0; j<l2; j++) {
            for( i=0; i<l1; i++) {
45                score = mat[ matidx( i, j, l1, l2)];
                if( score > maxscore) {
                    maxscore = score;
                    max_i      = i;
                    max_j      = j;
50                }
            }
        }
        if( ! firstscore && maxscore >= param->min_score) firstscore
55    = maxscore;

        push( stack, setpair( max_i, max_j));
        thescore = mat[ matidx( max_i, max_j, l1, l2)];
        mat[ matidx( max_i, max_j, l1, l2)] = 1;

```

```

/* Trace back */
k=0;
while( stack->sp > 0 ) {
5   pair = pop( stack);
   if(0){ printf( "pop      : " ); print_pair( pair);}
   i = pair.i;
   j = pair.j;
   if( path[ matidx( i, j, 11, 12)] == 1 ) {
10  if( i>0 && mat[ matidx( i-1, j, 11, 12)] > 0 ) {
       push( stack, setpair( i-1, j));
       }
       qseq[k] = seq1[i];
       mseq[k] = ' ';
15      tseq[k] = '-';
       k++;
       }
       else if( path[ matidx( i, j, 11, 12)] == 2 ) {
20      if( j>0 && mat[ matidx( i, j-1, 11, 12)] > 0 ) {
           push( stack, setpair( i, j-1));
           }
           qseq[k] = '-';
           mseq[k] = ' ';
           tseq[k] = seq2[j];
25          k++;
           }
           else if( path[ matidx( i, j, 11, 12)] == 0 ) {
               if( i>0 && j>0 && mat[ matidx( i-1, j-1, 11, 12)] > 0 ) {
30                  push( stack, setpair( i-1, j-1));
                  }
                  qseq[k] = seq1[i];
                  tseq[k] = seq2[j];
                  mseq[k] = match_sym( tseq[k], qseq[k], param-
35  >score_rec);
                  k++;
                  if(0){printf( " ( ) %d %d\n", j, i );
                      }
                  }
40      if( thescore >= param->min_score ) {
           n_hyp++;
           if( n_hyp <= param->max_res) {
               qseq[k] = 0;
               mseq[k] = 0;
45              tseq[k] = 0;
               outstr[0] = reverse( qseq);
               outstr[1] = reverse( mseq);
               outstr[2] = reverse( tseq);
               if( outfile != NULL ) {
50                  fprintf( outfile, "Score= %d\n", thescore);
                  print_lines( outfile, outstr, 3);
                  }
               }
           }
55      annot = empty_str( 11);
      free( annot);
}

```

```

    free( mt);
    free( qseq);
    free( mseq);
    free( tseq);
5   free_stack( stack);
    return firstscore;
}

10 void out_align_fill( param_type *param, int l1, int l2, int
    *idx_seq1,
        int *idx_seq2, int *mat, int *path, char *seq1, char *seq2){
    int i, j, k,    max_path;
    int score, max_score;
15   int h_size = 3;
    int gap_score;
    int match_factor;

    /*
20     0 = match or mismatch, go diagonally
        1 = insertion in query, go horizontally
        2 = gap in query, go vertically
    */

25   for( j=0; j<l2; j++) {
        for( i=0; i<l1; i++) {
            if(!printf( "%d %d\n", i, j));

            if( path[ matidx( i, j, l1, l2)] == -1 )
30             gap_score = param->score_rec->gap_cont;
            else
                gap_score = param->score_rec->gap_start;
            max_score = mat[ matidx( i, j, l1, l2)] -
gap_score;
35             max_path= -1;

            if( path[ matidx( i, j+1, l1, l2)] == -2 )
                gap_score = param->score_rec->gap_cont;
            else
40             gap_score = param->score_rec->gap_start;
            score = mat[ matidx( i, j+1, l1, l2)] - gap_score;
            if( score > max_score) { max_score = score; max_path =
-2; }

45             if( i-j <= h_size ) {
                score = param->score_rec->loop_score;
            }
            else {
                if( path[ matidx( i-1, j+1, l1, l2)] == 0 )
50                 match_factor = 1;
                else
                    match_factor = param->score_rec->match_cont_factor;
                score = mat[ matidx( i-1, j+1, l1, l2)] +
                    match_factor *
55                 param->score_rec->mat[ idx_seq1[i] *
                    param->score_rec->alph_len + idx_seq2[j]];
            }
        }
    }
}

```



```

        if(0) {
            printf( "%d %d %d %d    %d %d %d\n", i,    j, idx_seq1[i],
idx_seq2[j],
                param->score_rec->alph_len,
5          param->score_rec->mat[ idx_seq1[i] * param->score_rec-
>alph_len +
                idx_seq2[j]],
                idx_seq1[i] * param->score_rec->alph_len + idx_seq2[j]);
        }
10      if( score    > max_score) { max_score = score; max_path =
0;}
        for( k=j+1; k<i; k++) {
            score =    mat[ matidx( i, k+1, 11, 12)] + mat[ matidx( k,
15  j, 11, 12)]
            - param->score_rec->gap_start;
            if( score > max_score) { max_score = score; max_path = k;}
            if(0)printf( "%d %d %d %d %d\n", i, k, k, j, score);
        }
20      if( max_score > 0 ) {
            mat[ matidx( i, j, 11, 12)] = max_score;
        }
        else {
25      mat[ matidx( i, j, 11, 12)] = 0;
        }
        path[ matidx( i, j, 11, 12)] = max_path;
        if(0) {
            printf("%c%c  %3d %3d  score: %3d  path: %3d\n", seq1[i],
30  seq2[j],
                i, j, max_score, max_path);
        }
    }
35  }

void align_fill( param_type *param, int l1, int l2, int
*idx_seq1,
40      int *idx_seq2, int *mat, int *path, char *seq1, char *seq2){
    int i, j;
    char c1[2], c2[2];
    int i1, i2;
    int sm, si, sd, max_s;
45      int dir;

    if( verbose )
        printf( "%d    %d\n", l1, l2);

50      /*
        0 = match or mismatch, go diagonally
        1 = insertion in query, go horizontally
        2 = gap in    query, go vertically
        */
55      for( j=0; j<l2; j++) {
            c2[0] = seq2[j];
            c2[1] = 0;

```

```

i2 = strstr( param->score_rec->alph, c2);
for( i=0; i<11; i++) {
    c1[0] = seq1[i];
    c1[1] = 0;
5     i1 = strstr( param->score_rec->alph, c1);

    if( i > 0 && j > 0) {
10     sm = mat[(j-1)*11+(i-1)];
    }
    else {
        if( i > 0 ) {
            sm = param->score_rec->gap_start +
15             (i-1) * param->score_rec->gap_cont;
        }
        else if( j > 0 ) {
            sm = param->score_rec->gap_start +
                (j-1) * param->score_rec->gap_cont;
        }
20     sm = 0;
    }
    sm += param->score_rec->mat[ i2*param->score_rec-
>alph_len+i1];

25     /*if( (i > 0) && (path[j*11+(i-1)] > 0)) { */
    if( i > 0 ) {
        si = mat[j*11+(i-1)];
    }
    else {
30     si = 0;
    }
    if( (i>0 && path[j*11+(i-1)] > 0) || ( j>0 && path[(j-
1)*11+i] > 0)) {
35     si -= param->score_rec->gap_cont;
    }
    else {
        si -= param->score_rec->gap_start;
    }

40     /*if( (j > 0) && (path[(j-1)*11+i] > 0)) { */
    if( j > 0 ) {
        sd = mat[(j-1)*11+i];
    }
    else {
45     sd = 0;
    }
    sd -= param->score_rec->gap_cont;

    max_s = sm;
50     dir = 0;
    if( si > max_s) {
        max_s = si;
        dir = 1;
    }
55     if( sd > max_s) {
        max_s = sd;
        dir = 2;
    }

```

```

        mat[j*11+i] = MAX( max_s, 0);
        path[j*11+i] = dir;
        /*printf( "%s %s %d %d %d %d\n", c1, c2, i1, i2,
5  score, mat[j*11+i]);*/
    )
}

    if( 0) {
10  print_mat( seq1, seq2, mat);
        printf( "\n");
        print_mat( seq1, seq2, path);
        printf( "\n");
    }
15
    /* printf( "score= %d\n", mat[11*12-1]); */
}

20 int align( FILE *outfile, param_type *param, char *seq1, char
    *seq2){
    int l1, l2;
    int *mat, *path;
    char *mask_seq1, *mask_seq2;
25  int *idx_seq1, *idx_seq2;
    int score = INT_MAX;

    l1 = strlen( seq1);
30  l2 = strlen( seq2);
    mat = calloc( matsize( l1, l2), sizeof( int));
    path = calloc( matsize( l1, l2), sizeof( int));
    mask_seq1 = calloc( l1 + 1, sizeof( char));
    mask_seq2 = calloc( l2 + 1, sizeof( char));
35  strcpy( mask_seq1, seq1);
    strcpy( mask_seq2, seq2);

    if(0) {
40  print_mat( seq1, seq2, mat);
        printf( "\n");
        if(1) print_mat( seq1, seq2, path);
        printf( "\n");
    }
45
    align_fill( param, l1, l2, idx_seq1, idx_seq2, mat, path,
        mask_seq1, mask_seq2);
    score = align_local_traceback( outfile, param, l1, l2,
        idx_seq1, idx_seq2,
50  mask_seq1, mask_seq2, seq1, seq2, mat, path);
    free( mat);
    free( path);
    free( mask_seq1);
    free( mask_seq2);
55  return score;
}

```

```

/***** Nussinov - secondary structure prediction *****/

int nussinov_local_traceback( FILE *outfile, param_type *param,
5   int depth,
                                int l, int* idx_seq, char *mask_seq, char *seq,
                                int *mat, int *path){
    stack_type *stack;
    int i, j, k;
10   pair_type pair;
    char *annot;
    matrix_type *mt;
    char *mask;
    int score, maxscore, thescore=INT_MAX, max_i, max_j;
15   int maxpath;
    int overlap= 0;
    int n_hyp = 0;
    char *outstr[2];

20   stack = init_stack( STACK_CHUNK);

    mt = calloc( 1, sizeof( matrix_type));
    mt->d = mat;
    mt->l1 = 1;
25   mt->l2 = 1;

    annot = empty_str( 1);
    mask = empty_str( 1);

30   while( thescore >= param->min_score && n_hyp < param->max_res)
    {
        maxscore = INT_MIN;
        for( j=0; j<l; j++) {
15           if( mask[j] == ' ' ) {
                for( i=j; i<depth + j + 1 && i<l; i++) {
                    if( mask[i] == ' ' ) {
                        score = mat[ matidx( i, j, 1, 1)];
                        if( score > maxscore) {
40                             maxscore = score;
                                max_i = i;
                                max_j = j;
                            }
                        }
                    }
25           }
        }

        push( stack, setpair( max_i, max_j));
50       mask[ max_i] = 's';
        mask[ max_j] = 's';

        /* Trace back */

55       while( stack->sp > 0 ) {
            pair = pop( stack);
            if(0){ printf( "pop      : " ); print_pair( pair);}
            i = pair.i;

```

```

    j = pair.j;
    if( i > j ) {
    if( path[ matidx( i, j, 1, 1) ] == -1 ) {
    5      if( (overlap || mask[i-1] == ' ') &&
          mat[ matidx( i-1, j, 1, 1) ] > 0 ) {
          push( stack, setpair( i-1, j));
        }
        mask[i] = 'i';
    }
    10    else if( path[ matidx( i, j, 1, 1) ] == -2 ) {
        if( ( overlap || mask[j+1] == ' ') &&
          mat[ matidx( i, j+1, 1, 1) ] > 0 ) {
          push( stack, setpair( i, j+1));
        }
    15    mask[j] = 'd';
    }
    else if( path[ matidx( i, j, 1, 1) ] == 0 ) {
        if( ( overlap || (mask[i-1] == ' ' && mask[j+1] == ' '
    20    )) &&
          mat[ matidx( i-1, j+1, 1, 1) ] > 0 ) {
          push( stack, setpair( i-1, j+1));
        }
        if( param->score_rec->match_threshold <=
          param->score_rec->mat[ idx_seq[i] * param->score_rec-
    25    >alph_len + idx_seq[j]]) {
          annot[i] = ')';
          annot[j] = '(';
        }
        else {
    30          annot[i] = '.';
          annot[j] = '.';
        }
        mask_seq[i] = '-';
        mask_seq[j] = '-';
    35    mask[i] = 'm';
        mask[j] = 'm';
        if(0)printf( " ( ) %d %d\n", j,i );
    }
    else {
    40      k = path[ matidx( i, j, 1, 1)];
      push( stack, setpair( k, j));
      push( stack, setpair( i, k+1));
    }
    }
    45  }
    if(0){
        /* Trace forward */

    50    push( stack, setpair( max_i, max_j));

    while( stack->sp > 0 ) {
        pair = pop( stack);
        if(1){ printf( "pop      : " ); print_pair( pair);)
        i = pair.i;
    55    j = pair.j;
        if( i<depth + j && i<1-1 && j>0 ) {
            maxscore = 0;
            maxpath = -3;

```

```

    if( path[ matidx( i+1, j, 1, 1)] == -1 &&
        ( overlap || mask[i+1] == ' ' ) &&
        mat[ matidx( i+1, j, 1, 1)] >= maxscore ) {
5       maxscore = mat[ matidx( i+1, j, 1, 1)];
        maxpath = -1;
    }
    if( path[ matidx( i, j-1, 1, 1)] == -2 &&
        ( overlap || mask[j-1] == ' ' ) &&
10       mat[ matidx( i, j-1, 1, 1)] >= maxscore ) {
        maxscore = mat[ matidx( i, j-1, 1, 1)];
        maxpath = -2;
    }
    if( path[ matidx( i+1, j-1, 1, 1)] == 0 &&
        ( overlap || (mask[i+1] == ' ' && mask[j-1] ==
15 ' ')) &&
        mat[ matidx( i+1, j-1, 1, 1)] >= maxscore ) {
        maxscore = mat[ matidx( i+1, j-1, 1, 1)];
        maxpath = 0;
    }
20    if( path[ matidx( i+1, j-1, 1, 1)] > 0 &&
        mat[ matidx( i+1, j-1, 1, 1)] >= maxscore ) {
        maxscore = mat[ matidx( i+1, j-1, 1, 1)];
        maxpath = path[ matidx( i+1, j-1, 1, 1)];
    }
25    if( maxpath == -1 ) {
        push( stack, setpair( i+1, j));
        mask[i+1] = 'I';
    }
30    else if( maxpath == -2 ) {
        push( stack, setpair( i, j-1));
        mask[j-1] = 'D';
    }
    else if( maxpath == 0 ) {
35        push( stack, setpair( i+1, j-1));
        if( param->score_rec->match_threshold <=
            param->score_rec->mat[ idx_seq[i+1] * param-
>score_rec->alph_len
40            + idx_seq[j-1]]) {
            annot[i+1] = ')';
            annot[j-1] = '(';
        }
        else {
45            annot[i+1] = '.';
            annot[j-1] = '.';
        }
        mask[i+1] = 'M';
        mask[j-1] = 'M';
        if(0)printf( " ( ) %d %d\n", j, i );
50    }
    else if( maxpath > 0 ) {
        k = maxpath;
        push( stack, setpair( k, j-1));
        push( stack, setpair( i+1, k+1));
55    }
    }
}

```

```

thescore = mat[ matidx( max_i, max_j, 1, 1)];
if( thescore >= param->min_score ) {
    n_hyp++;
    if( n_hyp <= param->max_res ) {
5      if(0)printf( "%s\n", seq );
      if(0)printf( "%s\n", annot );
      outstr[0] = seq;
      outstr[1] = annot;
10     if( outfile != NULL ) {
        fprintf( outfile, "Score= %d\n", thescore);
        print_lines( outfile, outstr, 2);
    }
    }
15    free( annot);
    annot = empty_str( 1);
}

20    free( mt);
    free( annot);
    free( mask);
    free_stack( stack);
    return thescore;
25 }

char *nussinov_traceback( FILE *outfile, param_type *param,
                          int depth, int l, int* idx_seq,
30      char *seq, int *mat, int *path){
    stack_type *stack;
    int i, j, k, d;
    pair_type pair;
    char *annot;
35    matrix_type *mt;
    int gap_score1, gap_score2;

    stack = init_stack( STACK_CHUNK);

40    mt = calloc( 1, sizeof( matrix_type));
    mt->d = mat;
    mt->l1 = 1;
    mt->l2 = 1;

45    annot = empty_str( 1);

    for( d=depth; d<1; d++) {

50        push( stack, setpair( d, d-depth));

        while( stack->sp > 0 ) {
            pair = pop( stack);
            if(0){ printf( "pop      : " ); print_pair( pair);}
            i = pair.i;
55            j = pair.j;
            if( i > j ) {
                if( path[ matidx( i, j, 1, 1)] == -1 ) {
                    push( stack, setpair( i-1, j));
                }
            }
        }
    }
}

```

```

    }
    else if( path[ matidx( i, j, 1, 1) ] == -2 ) {
        push( stack, setpair( i, j+1));
    }
5   else if( path[ matidx( i, j, 1, 1) ] == 0 ) {
        if( param->score_rec->match_threshold <=
            param->score_rec->mat[ idx_seq[i] * param-
>score_rec->alph_len + idx_seq[j]] ) {
10            annot[i] = ')';
            annot[j] = '(';
        }
        else {
            annot[i] = '.';
            annot[j] = '.';
15        }
        if(0)printf( " ( ) %d %d\n", j, i );
        push( stack, setpair( i-1, j+1));
    }
    else {
20        k = path[ matidx( i, j, 1, 1)];
        push( stack, setpair( k, j));
        push( stack, setpair( i, k+1));
    }
    if(0){
25        if( path[ matidx( i-1, j, 1, 1) ] == -1 )
            gap_score1 = param->score_rec->gap_cont;
        else
            gap_score1 = param->score_rec->gap_start;
        if( path[ matidx( i, j+1, 1, 1) ] == -2 )
30            gap_score2 = param->score_rec->gap_cont;
        else
            gap_score2 = param->score_rec->gap_start;

        if( m( i-1, j, mt) ==
35            m( i, j, mt) - gap_score1 ) {
            if(0)printf( " mat: -1 %d %d\n", m( i-1, j, mt), m( i, j,
mt));
            push( stack, setpair( i-1, j));
        }
        else if( m( i, j+1, mt) ==
40            m( i, j, mt) - gap_score2 ) {
            if(0)printf( " mat: -2 %d %d\n", m( i-1, j, mt), m( i, j,
mt));
            push( stack, setpair( i, j+1));
        }
45        else if( (mat[ matidx( i-1, j+1, 1, 1) ] +
            param->score_rec->mat[ idx_seq[i] * param->score_rec-
>alph_len
50            + idx_seq[j]])
            ==
            mat[ matidx( i, j, 1, 1) ] ) {
            if(0)printf( " mat: 0 %d %d\n", m( i-1, j, mt), m( i, j,
mt));
            annot[i] = ')';
            annot[j] = '(';
55            if(0)printf( " ( ) %d %d\n", j, i );
            push( stack, setpair( i-1, j+1));
        }
    }

```



```

    else {
        for( k=j+1; k<i; k++) {
            if(0)printf( " %d %d %d %d\n", i, k, k, j);
            if( mat[ matidx( i, k+1, 1, 1)] + mat[ matidx( k,
5   j, 1, 1)] ==
                mat[ matidx( i, j, 1, 1)]) {
                if(0)printf( " mat: %d %d %d\n", k, m( i-1, j,
mt), m( i, j, mt));
                push( stack, setpair( k, j));
10             push( stack, setpair( i, k+1));
                break;
            }
        }
    }
15 }
}

fprintf( outfile, "Score = %d\n", mat[ matidx( d, d-depth,
20 1, 1)]);
printf( "%d>%s\n", d, seq );
printf( "%d>%s\n", d, annot );
free( annot);
annot = empty_str( 1);
25 }

free_stack( stack);
return annot;
30 }

void nussinov_fill( param_type *param, int depth, int l, int
*idx_seq,
35 int *mat, int *path, char *seq){
    int d;
    int i, j, k, max_path;
    int score, max_score;
    int h_size = 3;
40 int gap_score;
    int match_factor;

    for( d=0; d<depth; d++) {
        for( i=d+1; i<l; i++) {
45             if(0)printf( "%d %d %d\n", d, i, depth);
                j = i - 1 - d;

                if( path[ matidx( i-1, j, 1, 1)] == -1 )
                    gap_score = param->score_rec->gap_cont;
20             else
                    gap_score = param->score_rec->gap_start;
                    max_score = mat[ matidx( i-1, j, 1, 1)] -
gap_score;
                    max_path = -1;
55             if( path[ matidx( i, j+1, 1, 1)] == -2 )
                    gap_score = param->score_rec->gap_cont;
                else

```

```

gap_score = param->score_rec->gap_start;
score = mat[ matidx( i, j+1, 1, 1)] - gap_score;
if( score > max_score) { max_score = score; max_path =
-2;}
5
    if( i-j <= h_size ) {
        score = param->score_rec->loop_score;
    }
    else {
10    if( path[ matidx( i-1, j+1, 1, 1)] == 0 )
        match_factor = 1;
    else
        match_factor = param->score_rec->match_cont_factor;
    score = mat[ matidx( i-1, j+1, 1, 1)] +
15    match_factor *
        param->score_rec->mat[ idx_seq[i] ] - param->score_rec->
>alph_len +
        idx_seq[j]);
    }
20    if(0)printf( "%d %d %d %d %d %d %d\n", i, j,
idx_seq[i], idx_seq[j],
        param->score_rec->alph_len,
        param->score_rec->mat[ idx_seq[i] ] * param->score_rec-
>alph_len +
25    idx_seq[j]), idx_seq[i] * param->score_rec->alph_len +
idx_seq[j]);

    if( score > max_score) { max_score = score; max_path =
0;}
30    for( k=j+1; k<i; k++) {
        score = mat[ matidx( i, k+1, 1, 1)] + mat[ matidx( k, j,
1, 1)]
            - param->score_rec->gap_start;
        if( score > max_score) { max_score = score; max_path = k;}
35    if(0)printf( "%d %d %d %d %d\n", i, k, k, j, score);
    }

    if( max_score > 0 ) {
40    mat[ matidx( i, j, 1, 1)] = max_score;
    }
    else {
        mat[ matidx( i, j, 1, 1)] = 0;
    }
    path[ matidx( i, j, 1, 1)] = max_path;
45    if(0)printf( "%c%c %3d %3d score: %3d path: %3d\n",
seq[i], seq[j], i, j, max_score, max_path);
    }
}
50

int nussinov( FILE *outfile, param_type *param, char *seq){
    int l, i;
    int *mat, *path;
55    int *idx_seq;
    int score = INT_MAX;
    char *mask_seq;
    int max_score = 0;

```

```

int depth = param->depth;

1   = strlen( seq);
5   if( depth > 1-1 ) { depth = 1-1;}
    mat      = calloc( matsize( 1, 1), sizeof( int));
    path     = calloc( matsize( 1, 1), sizeof( int));
    mask_seq = calloc( 1+1, sizeof( char));
10  strcpy( mask_seq, seq);

    if(0) {
        print_mat( seq, seq, mat);
        printf( "\n");
15  if(0)print_mat( seq, seq, path);
        printf( "\n");
    }

    if( 0) {
20  idx_seq = seq2idx_seq( seq, param->score_rec);
        nussinov_fill( param, depth, 1, idx_seq, mat, path, seq);
        nussinov_traceback( outfile, param, depth, 1, idx_seq,
seq, mat, path);
    }
25  else {
        i = 0;
        while( score > param->min_score && i < param->max_res) {
            idx_seq = seq2idx_seq( mask_seq, param->score_rec);
            nussinov_fill( param, depth, 1, idx_seq, mat, path,
30  mask_seq);
            score = nussinov_local_traceback( outfile, param,
depth, 1,
idx_seq, mask_seq, seq, mat, path);
            i++;
35  if( score > max_score) max_score = score;
        }
        if( outfile != NULL ) {
            printf( "Mask:\n");
            if( score >= param->min_score)
40  print_lines( outfile, &mask_seq, 1);
        }
    }

45  free( mat);
    free( path);
    free( idx_seq);
    free( mask_seq);
    return max_score;
50 }

/***** OLIGO *****/
55 void calc_olig( int id, char *seq, char *rev_seq, param_type
*param){
    int score_self, score_target, score_struc;

```

```

        strcpy( rev_seq, seq);
        reverse( rev_seq);
        score_self = align( NULL, param, seq, rev_seq);
5       lower( rev_seq);
        comp( rev_seq);
        reverse( rev_seq);
        score_target = align( NULL, param, seq, rev_seq);
        score_struct = nussinov( NULL, param, seq);
10      printf( "%7d %s %7d %7d %7d\n", id, seq, score_self,
        score_target,
            score_struct);
    }

15
void allolig( FILE *outfile, int oligolen, int oligosample, int
spikefreq,
    param_type *param){
20     char alph[255] = "acgtACGT";
        long n;
        char *seq, *rev_seq;
        long j;

25     rev_seq = empty_str( oligolen);

        param->min_score = 2;
        param->max_res = 1;

30     if( oligosample ) {
        for( j=0; j<oligosample; j++) {
            seq = make_random_nmer( oligolen, spikefreq, alph);
            calc_olig( j, seq, rev_seq, param);
            free( seq);
35     }
        }
        else {
            n = n_nmer( oligolen, alph);
            for( j=0; j<n; j++) {
40                 seq = make_nmer( j, oligolen, alph);
                    calc_olig( j, seq, rev_seq, param);
                    free( seq);
            }
        }
45     free( rev_seq);
    }

50     /***** ARGUMENT PARSING AND INITIALIZATION *****/

void add_seq2seqs( char *seq, sequences_type *seqs){
    int new_size, i;
    if( seqs->nseq < seqs->maxnseq ) {
55         seqs->seqs[ seqs->nseq++] = seq;
    }
    else {
        new_size = (seqs->maxnseq + SEQS_CHUNK) * sizeof( char *);

```

```

    if( seqs->seqs = realloc( seqs->seqs, new_size)) {
        seqs->maxnseq = seqs->maxnseq + SEQS_CHUNK;
        for( i=seqs->nseq; i<seqs->maxnseq; i++)
            seqs->seqs[i] = NULL;
5       seqs->seqs[ seqs->nseq++] = seq;
    }
    else {
        die( "Failed to realloc seq array to %d bytes", new_size);
    }
10  }
    )

```

```

void read_fasta( FILE *seqfile, sequences_type *seqs, char
15  *alph){

```

```

    int c;
    char name[255];
    char comment[255];
    char s[255];
20  int seq_flag = 0;
    dynamic_str_type *seq = NULL;

    while( (c = fgetc( seqfile)) != EOF){
        if( c == '>') {
25         if( seq != NULL && seq->len > 0 ) {
            add_seq2seqs( seq->s, seqs);
            free( seq);
            seq = new_dynamic_str_type();
        }
        fgets( s, 254, seqfile);
        sscanf( s, "%s%s", name, comment);
        seq_flag = 1;
    }
    else if( seq_flag == 1 ) {
35         if( strchr( alph, c) != NULL ) {
            seq = addchar( seq, c);
        }
    }
40  if( seq != NULL && seq->len > 0 )
        add_seq2seqs( seq->s, seqs);

    free( seq);
45  }

```

```

param_type *init(){
    char a[255];
    int i, j;
50  param_type *param;
    int *mat;
#define ALPHLEN 9
    int mathyp_init[ALPHLEN][ALPHLEN]= {
55         { -3, -3, -3, 3, -5, -5, 5, -99},
            { -3, -3, 5, -3, -5, -5, 9, -99},
            { -3, 5, -3, 1, -5, 9, -5, 2, -99},
            { 3, -3, 1, -3, 5, -5, 2, -5, -99},
            { -5, -5, -5, 5, -8, -8, 8, -99},

```

```

        { -5, -5, 9, -5, -8, -8, 14, -8, -99},
        { -5, 9, -5, 2, -8, 14, -8, 4, -99},
        { 5, -5, 2, -5, 8, -8, 4, -8, -99},
        {-99,-99,-99,-99,-99,-99,-99,-99,-99}
5      );

    param = calloc( 1, sizeof( param_type));

    /* alph */
10   strcpy( a, "acgtACGT-");
    param->alph = calloc( strlen( a), sizeof( char));
    strcpy( param->alph, a);

    /* seqs */
15   param->seqs = calloc( 1, sizeof( sequences_type));
    param->seqs->nseq = 0;
    param->seqs->maxnseq = SEQS_CHUNK;
    param->seqs->seqs = calloc( param->seqs->maxnseq, sizeof(
20   char *));
    param->seqs2 = calloc( 1, sizeof( sequences_type));
    for( i=0; i<param->seqs->maxnseq; i++)
        param->seqs->seqs[i] = NULL;
    param->seqs2->nseq = 0;
    param->seqs2->maxnseq = SEQS_CHUNK;
25   param->seqs2->seqs = calloc( param->seqs2->maxnseq, sizeof(
    char *));
    for( i=0; i<param->seqs2->maxnseq; i++)
        param->seqs2->seqs[i] = NULL;

30   /* hybridization matrix */

    mat = calloc( ALPHLEN * ALPHLEN, sizeof( int));
    for( i=0; i<ALPHLEN; i++){
        for( j=0; j<ALPHLEN; j++){
35         mat[i*ALPHLEN+j] = mathyp_init[i][j];
        }
    }

    /* defaults */
40   param->depth = INT_MAX;
    param->min_score = 20;
    param->max_res = INT_MAX;

45   param->score_rec = init_score_rec_type(
        14, 7,
        -40, 2, 1,
        8, 3, 1,
        param->alph, mat);
50   validate_score_rec_type( param->score_rec, -200, 200);

    return param;
55 }

/***** MAIN *****/

```

```

int main (int argc, char* argv[]) {
    int      i;
    char      *rev_seq;
5    param_type *param;
    int      algo          = 0;

    char      *options;
    int      next_option;
10    char*      outfn      = NULL;
    char*      infn        = NULL;
    FILE      *outfile, *seqfile;
    int      othermethod = 0;
    int      oligolen = 7;
15    int      oligosample = 0;
    int      spikefreq    = 0;

    const char* const short_options =
    "hvrwayto:i:j:s:z:d:n:p:m:l:e:f:";
20    const struct option long_options[] = {
        { "help",          0, NULL, 'h' },
        { "verbose",       0, NULL, 'v' },
        { "version",       0, NULL, 'x' },
        { "secondary_structure", 0, NULL, 'u' },
25    { "self_annealing",   0, NULL, 'a' },
        { "hybridization",  0, NULL, 'y' },
        { "tm",            0, NULL, 't' },
        { "output",        1, NULL, 'o' },
        { "input",         1, NULL, 'i' },
30    { "seq",             1, NULL, 's' },
        { "seq2",          1, NULL, 'z' },
        { "depth",         1, NULL, 'd' },
        { "min_score",     1, NULL, 'n' },
        { "max_res",       1, NULL, 'p' },
35    { "matrix",          1, NULL, 'm' },
        { "allolig",       1, NULL, 'l' },
        { "sample",        1, NULL, 'e' },
        { "spikefreq",     1, NULL, 'f' },
40    { NULL,              0, NULL, 0 } /* Required at end of
        array. */
    };

    /* mtrace(); debug memory leaks under linux */

45    param = init();

    program_name = argv[0];

    do {
50        next_option = getopt_long (argc, argv, short_options,
            long_options, NULL);
        switch (next_option)
        {
            case 'h': /* -h or --help */
55            usage ("");

            case 'i': /* -i or --input */
                infn = optarg;

```

```

    seqfile =    sfopen(    infn, "r");
    read_fasta( seqfile, param->seqs,    param->alph);
    fclose( seqfile);
    break;
5
    case 'j': /* -j or --input2 */
        infn = optarg;
        seqfile =    sfopen(    infn, "r");
        read_fasta( seqfile, param->seqs2, param->alph);
10        fclose( seqfile);
        break;

    case 'o': /* -o or --output */
        outfn = optarg;
15        break;

    case 'm': /* -m or --matrix*/
        options =    optarg;
        read_score_rec_type( options, param->score_rec);
20        break;

    case 's': /* -s or --seq*/
        options =    optarg;
        add_seq2seqs( options, param->seqs);
25        break;

    case 'z': /* -z or --seq2*/
        options =    optarg;
        add_seq2seqs( options, param->seqs2);
30        break;

    case 'd': /* -d or --depth*/
        options =    optarg;
        param->depth = atoi( options);
35        break;

    case 'p': /* -p or --max_res*/
        options =    optarg;
        param->max_res = atoi( options);
40        break;

    case 'n': /* -n or --min_score*/
        options =    optarg;
        param->min_score = atoi( options);
45        break;

    case 'v': /* -v or --verbose */
        verbose += 1;
50        break;

    case 'r': /* -r or --version */
        printf( "dyp version %s\n", THEVERSION);
        break;

55    case 'u': /* -s or --secondary_structure */
        algo = algo | SECONDARY_STRUCTURE;
        break;

```



```

    case 'a': /* -s or --self_anealign */
        algo = algo | SELF_ANEALING;
        break;

5   case 'y': /* -s or --self_anealign */
        algo = algo | HYBRIDIZATION;
        break;

    case 't': /* -t or --tm */
10      algo = algo | TM;
        break;

    case 'l': /* -l or --allolig */
        options = optarg;
15      oligolen = atoi( options);
        othermethod = -1;
        break;

    case 'e': /* -p or --sample */
20      options = optarg;
        oligosample = atoi( options);
        break;

    case 'f': /* -f or --spikefreq */
25      options = optarg;
        spikefreq = atoi( options);
        break;

    case '?': /* The user specified an invalid option. */
30      usage ("Sorry, one of the options was not
        recogniced");

    case -1: /* Done with options. */
35      break;

    default: /* Something else: unexpected. */
        abort ();
    }
40 while (next_option != -1);

    if( outfn == NULL )
        outfile = stdout;
45     else
        outfile = sfopen( outfn, "w");

    if( verbose)
50      print_score_rec_type( param->score_rec);

    if( othermethod) {
        allolig( outfile, oligolen, oligosample, spikefreq, param);
    }
55     else {
        if( !algo )
            usage( "please select a method: u, a or y");
    }

```

```

    for( i=0; i<param->seqs->nseq; i++) {
        fprintf( outfile, ">sequence_%03d\n", i+1);
        if( algo & SECONDARY_STRUCTURE )
            nussinov( outfile, param, param->seqs->seqs[i]);
5         if( algo & SELF_ANEALING ) {
            rev_seq = calloc( strlen( param->seqs->seqs[i]), sizeof(
char));
            strcpy( rev_seq, param->seqs->seqs[i]);
            reverse( rev_seq);
10         if( param->min_score < 2 ) {
            param->min_score = 2;
        }
        align( outfile, param, param->seqs->seqs[i], rev_seq);
        free( rev_seq);
15     }
    if( algo & TM ) {
        rev_seq = calloc( strlen( param->seqs->seqs[i]), sizeof(
char));
        strcpy( rev_seq, param->seqs->seqs[i]);
20         lower( rev_seq);
        comp( rev_seq);
        if( param->min_score < 2 ) {
            param->min_score = 2;
        }
25     align( outfile, param, param->seqs->seqs[i], rev_seq);
        free( rev_seq);
    }
    if( algo & HYBRIDIZATION ) {
        if( param->seqs2->seqs[i] == NULL)
30         die( "Sequence 2 is missing");
        rev_seq = calloc( strlen( param->seqs2->seqs[i]), sizeof(
char));
        strcpy( rev_seq, param->seqs2->seqs[i]);
        reverse( rev_seq);
35     align( outfile, param, param->seqs->seqs[i], rev_seq);
        free( rev_seq);
    }
}
40 fclose( outfile);

    free_param_type( param);

    return 0;
45 }

```

Figure 28 C

1

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

```

<PARAMETER>
oligo_length      30
blastdb           /home/probe/database/bf/c_elegans
mindist          30
5
#blast_param      -e 50 -F F -a 4
blast_param       wordlen  11
blast_param       strand  both strands
blast_param       expect   5
10 blast_param      nproc    2
blast_param       filter   F

cfreq             3
cphase            0
15 end_spike_len    0
dnaconc           2000
saltconc          115
max_noligo        10
maxhits           8
20 min_score       35

#perceptron parameters
oligod_param      max_match      cutoff      30
25 oligod_param      max_match      squash_dx    5
oligod_param      max_match      squash_dy    0.9
oligod_param      max_match      weight       1
oligod_param      max_stretch    cutoff       20
oligod_param      max_stretch    squash_dx    2
30 oligod_param      max_stretch    squash_dy    0.9
oligod_param      max_stretch    weight       1
oligod_param      self_hyp       cutoff       25
oligod_param      self_hyp       squash_dx    10
oligod_param      self_hyp       squash_dy    0.9
35 oligod_param      self_hyp       weight       1
oligod_param      self_match     cutoff       50
oligod_param      self_match     squash_dx    10
oligod_param      self_match     squash_dy    0.9
oligod_param      self_match     weight       0
40 oligod_param      tm_min       cutoff       40
oligod_param      tm_min       squash_dx    2
oligod_param      tm_min       squash_dy    0.9
oligod_param      tm_min       weight       1
oligod_param      tm_max       cutoff       95
45 oligod_param      tm_max       squash_dx    2
oligod_param      tm_max       squash_dy    0.9
oligod_param      tm_max       weight       1
oligod_param      tm           cutoff       1.7
oligod_param      tm           squash_dx    0.1
50 oligod_param      tm           squash_dy    0.9
oligod_param      tm           weight       5
oligod_param      tm_dan_min     cutoff       50
oligod_param      tm_dan_min     squash_dx    2
oligod_param      tm_dan_min     squash_dy    0.9
55 oligod_param      tm_dan_min     weight       0
oligod_param      tm_dan_max     cutoff       95
oligod_param      tm_dan_max     squash_dx    2
oligod_param      tm_dan_max     squash_dy    0.9

```

	oligod_param	tm_dan_max	weight	0
	oligod_param	tm_dan	cutoff	1
	oligod_param	tm_dan	squash_dx	0.1
	oligod_param	tm_dan	squash_dy	0.9
5	oligod_param	tm_dan	weight	0
	oligod_param	cg	weight	0
	oligod_param	ggg	weight	0
	oligod_param	hit_score	weight	0
10	</PARAMETER>			

Figure 28D

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

1

```
// $Id: Tm_predictor.java,v 1.6 2002/10/17 14:12:50 jgk Exp $
```

```
5 package tmpred;

import java.io.*;
import java.util.*;
import java.lang.*;
10 import java.text.*;

// Compile with:
//      /usr/java1.2/bin/javac      -classpath      ../usr/java1.2/
tmpred/Tm_predictor.java

15 // GCC compile:
//      gcj      --main=tmpred.Tm_predictor      tmpred/Tm_predictor.java
tmpred/Tm_thermodynamic_model.java      tmpred/BadOligoException.java      -o
Tm_predictor

20 public abstract class Tm_predictor {

    public abstract double calcTm( String sequence ) throws
BadOligoException;

25     public abstract String lastSequence();

    public abstract Double lastTm();

30     public abstract String getParameter( String paramName );

    public abstract void setParameter( String paramName, String value
);

35     public abstract Enumeration getParameterNames();

    public static String getVersion(){ return "$Id: Tm_predictor.java,v
1.6 2002/10/17 14:12:50 jgk Exp $"; }

40     /** Singleton instance. */
    private static Tm_predictor instance = null;

    /** Return a singleton instance of the default implementation.
    *
45     * @author Kim Haagenzen
    */
    public static Tm_predictor getTm_predictor() {
        if ( instance == null ) {
            instance = getInstance();
        }
    }
}
```

```

    }
    return instance;
}

5  /** Return an instance of the default implementation.
    */
    public static Tm_predictor getInstance(){
        // Default model (for now..)
        return new Tm_thermodynamic_model();
10 }

    // For demo purposes
    public static void main (String args[]) {
        Tm_predictor tmmodel=Tm_predictor.getInstance();
15 Enumeration pnames=tmmodel.getParameterNames();
        String paramName;

        // tmmodel.setParameter("debug","true");

20 System.out.println("Version: "+tmmodel.getVersion());
        System.out.println();

        System.out.println("Constants:");
        while(pnames.hasMoreElements()){
25     paramName=(String)pnames.nextElement();

        System.out.println(paramName+"="+tmmodel.getParameter(paramName));
        }

30 System.out.println();
        System.out.println("Results:");
        System.out.println("Input Sequence="+args[0]);

        try{
35     System.out.println("Tm="+tmmodel.calcTm(args[0]));
        System.out.println("Translated                               Sequence:
"+tmmodel.lastSequence());
        }
        catch(BadOligoException expt){
40     System.out.println("Error: "+expt.getMessage());
        System.out.println("Tm= "+tmmodel.lastTm());
        System.out.println("Translated                               Sequence:
"+tmmodel.lastSequence());
        }
45 }
}

```

Figure 28E

19 MAJ 2003

Modtaget

1

```
package tmpred;

5  import java.io.*;
    import java.util.*;
    import java.lang.*;
    import java.text.*;

10  class Tm_param {
        public Hashtable table_deltaS_init, table_deltaH_init, table_H_nn,
        table_S_nn, table_H_mono;
    };

15  class Tm_thermodynamic_model extends Tm_predictor {
        private Tm_param modparam[];

        private Hashtable table_LD_translate, table_ws_translate,
20  table_ten_translate;

        private Properties parameterProperties;
        private String sequence; // Last translated sequence
        private Double tm; // Last calculated Tm

25  // Get methods
        public String lastSequence(){ return sequence; }
        public Double lastTm(){ return tm; }

        // Property methods
30  public String getParameter(String paramName){
            return (String)parameterProperties.get(paramName);
        }
        public void setParameter(String paramName, String value){
            parameterProperties.put(paramName, value);
35  }
        public Enumeration getParameterNames(){
            return parameterProperties.propertyNames();
        }

40  // Utility functions:
        private double hget(Hashtable h, String key){
            // get double value from hashtable, zero if it wasn't there.

            if(key!=null){
45  Double val=(Double)h.get(key);
                if(val != null){
                    return val.doubleValue();
                } else {
                    return 0.0;
                }
            }
        }
    }
}
```

```

    }
    }

    return 0.0;
5    }

    private void hcount(Hashtable h, String key){
        // Count entry in h(key) up one, if it exists.

10    if(key!=null){
        Double val=(Double)h.get(key);

        if(val!=null){
            h.put(key, new Double(val.doubleValue()+1));
15        } else {
            h.put(key, new Double(1));
        }
    }
}

20    private String trans_LD(String seq){
        int i;
        String result="";

25        for(i=0; i<seq.length(); i++){
            result=result
            (String)table_LD_translate.get(seq.substring(i,i+1));
        }

30        return result;
    }

    private String trans_ws(String seq){
        int i;
35        String result="";

        for(i=0; i<seq.length(); i++){
            result=result
            (String)table_ws_translate.get(seq.substring(i,i+1));
40        }

        return result;
    }

45    private String findReplace(String inpsq, String find, String
    replace){
        String subseq;
        String seq=inpsq;
        int i;

```



```

    for(i=0; i<seq.length()-find.length()+1; i++){
        subseq=seq.substring(i,i+find.length());
        if(subseq.equals(find)){
5
            seq=seq.substring(0,i)+replace+seq.substring(i+find.length());
        }
    }

10
    return seq;
}

private void checkSequence(String sequence) throws
----- BadOligoException {
15
    int goodlength=0;
    int i;
    char mon;
    String badletters="";

20
    for(i=0; i<sequence.length(); i++){
        mon=sequence.charAt(i);

        switch(mon){
25
            case 'a':
            case 'g':
            case 'c':
            case 't':
            case 'A':
            case 'G':
30
            case 'C':
            case 'T': goodlength++; break;
            default:
                if(badletters==""){
                    badletters=""+"mon;
35
                } else {
                    badletters=badletters+" "+mon;
                }
            }
        }

40
        if (sequence.length()!=goodlength){
            this.tm=null;
            // this.sequence="";
            throw new BadOligoException("Wrong monomer(s) in sequence:
45
            "+badletters);
        }
    }

    Tm_thermodynamic_model() {

```

```

parameterProperties=new Properties();
parameterProperties.put("debug","false"); // Set to true if
debugging
parameterProperties.put("oligo_conc","0.000002"); // =2
5 micromolar, this is the total strand concentration, when the two
complements are in 1 uM each
parameterProperties.put("salt_conc","0.115"); // =115mM, Exiqon
most common concentration

10 sequence=null;
tm=null;

// Translate tables:
15 table_LD_translate=new Hashtable();
table_LD_translate.put("a", "D");
table_LD_translate.put("g", "D");
table_LD_translate.put("c", "D");
table_LD_translate.put("t", "D");
table_LD_translate.put("A", "L");
20 table_LD_translate.put("G", "L");
table_LD_translate.put("C", "L");
table_LD_translate.put("T", "L");

table_ws_translate=new Hashtable();
25 table_ws_translate.put("a", "w");
table_ws_translate.put("g", "s");
table_ws_translate.put("c", "s");
table_ws_translate.put("t", "w");
table_ws_translate.put("A", "W");
30 table_ws_translate.put("G", "S");
table_ws_translate.put("C", "S");
table_ws_translate.put("T", "W");

table_ten_translate=new Hashtable();
35 table_ten_translate.put("aa", "-aa");
table_ten_translate.put("tt", "-aa");
table_ten_translate.put("at", "-at");
table_ten_translate.put("ta", "-ta");
table_ten_translate.put("ca", "-ca");
40 table_ten_translate.put("tg", "-ca");
table_ten_translate.put("gt", "-gt");
table_ten_translate.put("ac", "-gt");
table_ten_translate.put("ct", "-ct");
table_ten_translate.put("ag", "-ct");
45 table_ten_translate.put("ga", "-ga");
table_ten_translate.put("tc", "-ga");
table_ten_translate.put("cg", "-cg");
table_ten_translate.put("gc", "-gc");
table_ten_translate.put("gg", "-gg");

```

```

table_ten_translate.put("cc", "-gg");

//
// Model Parameters
5 //

modparam=new Tm_param[7];

// =====
10 // Parameterfil: /home/jgk/projects/tmpred/param/param.tmp

modparam[0]=new Tm_param();

modparam[0].table_deltaS_init=new Hashtable();
15 modparam[0].table_deltaH_init=new Hashtable();
modparam[0].table_H_nm=new Hashtable();
modparam[0].table_S_nm=new Hashtable();
modparam[0].table_H_mono=new Hashtable();

20 // deltaS_init
modparam[0].table_deltaS_init.put("A", new
Double(2.63578811043473));
modparam[0].table_deltaS_init.put("C", new Double(-
4.60854049293735));
25 modparam[0].table_deltaS_init.put("G", new Double(-
4.56687267801977));
modparam[0].table_deltaS_init.put("T", new
Double(2.7258379273728));
modparam[0].table_deltaS_init.put("s", new Double(-2.8));
30 modparam[0].table_deltaS_init.put("w", new Double(4.1));

// deltaH_init
modparam[0].table_deltaH_init.put("A", new
Double(1.41884851398012));
35 modparam[0].table_deltaH_init.put("C", new Double(-
2.01835247107257));
modparam[0].table_deltaH_init.put("G", new Double(-
0.740703588206669));
modparam[0].table_deltaH_init.put("T", new
40 Double(1.19531513535407));
modparam[0].table_deltaH_init.put("s", new Double(0.1));
modparam[0].table_deltaH_init.put("w", new Double(2.3));

// H_mono
45 modparam[0].table_H_mono.put("A", new Double(-15.5177140147424));
modparam[0].table_H_mono.put("C", new Double(-17.840200113071));
modparam[0].table_H_mono.put("G", new Double(-18.2150417538429));
modparam[0].table_H_mono.put("T", new Double(-16.4741049031613));
modparam[0].table_H_mono.put("s", new Double(-9.36));

```

```

modparam[0].table_H_mono.put("w", new Double(-7.35));

// H_nn
modparam[0].table_H_nn.put("-aa", new Double(-0.55));
5 modparam[0].table_H_nn.put("-at", new Double(0.15));
modparam[0].table_H_nn.put("-ca", new Double(-0.145));
modparam[0].table_H_nn.put("-cg", new Double(-1.24));
modparam[0].table_H_nn.put("-ct", new Double(0.555));
modparam[0].table_H_nn.put("-ga", new Double(0.155));
10 modparam[0].table_H_nn.put("-gc", new Double(-0.44));
modparam[0].table_H_nn.put("-gg", new Double(1.36));
modparam[0].table_H_nn.put("-gt", new Double(-0.045));
modparam[0].table_H_nn.put("-ta", new Double(0.15));
modparam[0].table_H_nn.put("SS", new Double(-0.130528857633394));
15 modparam[0].table_H_nn.put("SW", new Double(-1.0637157585591));
modparam[0].table_H_nn.put("Ss", new Double(-0.357926846466049));
modparam[0].table_H_nn.put("Sw", new Double(0.379281367679472));
modparam[0].table_H_nn.put("WS", new Double(-0.405577627528957));
modparam[0].table_H_nn.put("WW", new Double(-0.726103393389269));
20 modparam[0].table_H_nn.put("Ws", new Double(1.04679214548348));
modparam[0].table_H_nn.put("Ww", new Double(0.0429919627021588));
modparam[0].table_H_nn.put("sS", new Double(0.543872586888427));
modparam[0].table_H_nn.put("sW", new Double(0.575455160673995));
modparam[0].table_H_nn.put("ws", new Double(0.865853023978453));
25 modparam[0].table_H_nn.put("ww", new Double(0.789896125879916));

// S_nn
modparam[0].table_S_nn.put("-aa", new Double(-22.2));
modparam[0].table_S_nn.put("-at", new Double(-20.4));
30 modparam[0].table_S_nn.put("-ca", new Double(-22.7));
modparam[0].table_S_nn.put("-cg", new Double(-27.2));
modparam[0].table_S_nn.put("-ct", new Double(-21));
modparam[0].table_S_nn.put("-ga", new Double(-22.2));
modparam[0].table_S_nn.put("-gc", new Double(-24.4));
35 modparam[0].table_S_nn.put("-gg", new Double(-19.9));
modparam[0].table_S_nn.put("-gt", new Double(-22.4));
modparam[0].table_S_nn.put("-ta", new Double(-21.3));
modparam[0].table_S_nn.put("DL", new Double(-32.6588469014549));
modparam[0].table_S_nn.put("LD", new Double(-31.5601899937808));
40 modparam[0].table_S_nn.put("LL", new Double(-46.8237053519389));

// =====
//
45 // Parameterfil:
/home/jgk/projects/tmpred/param/4.4.0.3.48.myhomo.ttal13

modparam[1]=new Tm_param();

modparam[1].table_deltaS_init=new Hashtable();

```

```

modparam[1].table_deltaH_init=new Hashtable();
modparam[1].table_H_nn=new Hashtable();
modparam[1].table_S_nn=new Hashtable();
modparam[1].table_H_mono=new Hashtable();

5      // deltaS_init
      modparam[1].table_deltaS_init.put("A",          new      Double(-
0.417636505370175));
      modparam[1].table_deltaS_init.put("C",          new      Double(-
10  0.336962265530943));
      modparam[1].table_deltaS_init.put("G",          new      Double(0.158722707567369));
      modparam[1].table_deltaS_init.put("T",          new      Double(-
0.208276246703806));
15      modparam[1].table_deltaS_init.put("s",          new      Double(-
0.372560237115056));
      modparam[1].table_deltaS_init.put("w",          new      Double(-
0.369231842748097));

20      // deltaH_init
      modparam[1].table_deltaH_init.put("A",          new      Double(0.782799015317204));
      modparam[1].table_deltaH_init.put("C",          new      Double(0.505143958976093));
25      modparam[1].table_deltaH_init.put("G",          new      Double(-
0.278435409858967));
      modparam[1].table_deltaH_init.put("T",          new      Double(0.435388592178298));
      modparam[1].table_deltaH_init.put("s",          new      Double(0.674274914671847));
30      modparam[1].table_deltaH_init.put("w",          new      Double(0.737204216419899));

      // H_mono

35      // H_nn
      modparam[1].table_H_nn.put("-aa", new Double(-7.48976721304652));
      modparam[1].table_H_nn.put("-at", new Double(-7.32484387195421));
      modparam[1].table_H_nn.put("-ca", new Double(-7.87301403612122));
40      modparam[1].table_H_nn.put("-cg", new Double(-8.42170498069151));
      modparam[1].table_H_nn.put("-ct", new Double(-7.74869970168917));
      modparam[1].table_H_nn.put("-ga", new Double(-7.76121630019634));
      modparam[1].table_H_nn.put("-gc", new Double(-8.50240032966233));
      modparam[1].table_H_nn.put("-gg", new Double(-8.25010916492258));
45      modparam[1].table_H_nn.put("-gt", new Double(-7.93137471336983));
      modparam[1].table_H_nn.put("-ta", new Double(-7.19824578597169));
      modparam[1].table_H_nn.put("DL", new Double(-7.64614418904991));
      modparam[1].table_H_nn.put("LD", new Double(-8.33674742394922));
      modparam[1].table_H_nn.put("LL", new Double(-8.48734357858688));

```

```

// S_nn
modparam[1].table_S_nn.put("-aa", new Double(-21.0183983393758));
modparam[1].table_S_nn.put("-at", new Double(-20.9296841275317));
5 modparam[1].table_S_nn.put("-ca", new Double(-20.9035893027225));
modparam[1].table_S_nn.put("-cg", new Double(-20.6106325122583));
modparam[1].table_S_nn.put("-ct", new Double(-20.9057484730661));
modparam[1].table_S_nn.put("-ga", new Double(-20.9212322254303));
modparam[1].table_S_nn.put("-gc", new Double(-20.5901247977925));
10 modparam[1].table_S_nn.put("-gg", new Double(-20.7858543446338));
modparam[1].table_S_nn.put("-gt", new Double(-20.8689105244148));
modparam[1].table_S_nn.put("-ta", new Double(-20.9664930077991));
modparam[1].table_S_nn.put("AA", new Double(-21.5617306575376));
modparam[1].table_S_nn.put("AC", new Double(-20.3786877774106));
15 modparam[1].table_S_nn.put("AG", new Double(-19.7518872341502));
modparam[1].table_S_nn.put("AT", new Double(-21.2727546056293));
modparam[1].table_S_nn.put("Aa", new Double(-20.938973530993));
modparam[1].table_S_nn.put("Ac", new Double(-20.2351610010894));
modparam[1].table_S_nn.put("Ag", new Double(-19.4518499755888));
20 modparam[1].table_S_nn.put("At", new Double(-18.0784147699652));
modparam[1].table_S_nn.put("CA", new Double(-20.7498315076301));
modparam[1].table_S_nn.put("CC", new Double(-18.5922344843667));
modparam[1].table_S_nn.put("CG", new Double(-18.0631881798201));
modparam[1].table_S_nn.put("CT", new Double(-20.9602092321618));
25 modparam[1].table_S_nn.put("Ca", new Double(-19.34976501959));
modparam[1].table_S_nn.put("Cc", new Double(-20.0636169441132));
modparam[1].table_S_nn.put("Cg", new Double(-18.939750726347));
modparam[1].table_S_nn.put("Ct", new Double(-21.6068106111011));
modparam[1].table_S_nn.put("GA", new Double(-20.6285173614998));
30 modparam[1].table_S_nn.put("GC", new Double(-21.1393042416962));
modparam[1].table_S_nn.put("GG", new Double(-19.2292577000144));
modparam[1].table_S_nn.put("GT", new Double(-20.5200676805333));
modparam[1].table_S_nn.put("Ga", new Double(-18.7514389680752));
modparam[1].table_S_nn.put("Gc", new Double(-18.6326804946465));
35 modparam[1].table_S_nn.put("Gg", new Double(-19.4129295709103));
modparam[1].table_S_nn.put("Gt", new Double(-19.7139582131827));
modparam[1].table_S_nn.put("TA", new Double(-21.7225383328128));
modparam[1].table_S_nn.put("TC", new Double(-19.0166955249448));
modparam[1].table_S_nn.put("TG", new Double(-18.1432831834084));
40 modparam[1].table_S_nn.put("TT", new Double(-21.702655506544));
modparam[1].table_S_nn.put("Ta", new Double(-21.6061975277002));
modparam[1].table_S_nn.put("Tc", new Double(-21.5479504447966));
modparam[1].table_S_nn.put("Tg", new Double(-20.6436696433912));
modparam[1].table_S_nn.put("Tt", new Double(-21.6266985715776));
45 modparam[1].table_S_nn.put("AA", new Double(-20.5319311697475));
modparam[1].table_S_nn.put("aC", new Double(-19.025255715981));
modparam[1].table_S_nn.put("aG", new Double(-20.3463306063697));
modparam[1].table_S_nn.put("aT", new Double(-20.3507644210827));
modparam[1].table_S_nn.put("cA", new Double(-18.6568044948628));

```

```

modparam[1].table_S_nn.put("cC", new Double(-20.8609785312075));
modparam[1].table_S_nn.put("cG", new Double(-19.1534484621275));
modparam[1].table_S_nn.put("cT", new Double(-18.949763882851));
modparam[1].table_S_nn.put("gA", new Double(-19.9252395614033));
5 modparam[1].table_S_nn.put("gC", new Double(-19.5747039562859));
modparam[1].table_S_nn.put("gG", new Double(-19.4942636766488));
modparam[1].table_S_nn.put("gT", new Double(-19.9443842009211));
modparam[1].table_S_nn.put("tA", new Double(-22.5442702941593));
modparam[1].table_S_nn.put("tC", new Double(-19.7656230175562));
10 modparam[1].table_S_nn.put("tG", new Double(-20.6803075424075));
modparam[1].table_S_nn.put("tT", new Double(-20.9918041722088));

// =====
15 // Parameterfil:
/home/jgk/projects/tmpred/param/4.4.0.3.48.myhomo.ttugetre7

modparam[2]=new Tm_param();

20 modparam[2].table_deltaS_init=new Hashtable();
modparam[2].table_deltaH_init=new Hashtable();
modparam[2].table_H_nn=new Hashtable();
modparam[2].table_S_nn=new Hashtable();
modparam[2].table_H_mono=new Hashtable();

25 // deltaS_init
modparam[2].table_deltaS_init.put("A", new Double(-
0.407793407213917));
modparam[2].table_deltaS_init.put("C", new Double(-
30 0.213527372713559));
modparam[2].table_deltaS_init.put("G", new
Double(0.35848328509707));
modparam[2].table_deltaS_init.put("T", new
Double(0.203944086326224));
35 modparam[2].table_deltaS_init.put("S", new Double(-
0.372560237115056));
modparam[2].table_deltaS_init.put("w", new Double(-
0.369231842748097));

40 // deltaH_init
modparam[2].table_deltaH_init.put("A", new
Double(0.486242607856365));
modparam[2].table_deltaH_init.put("C", new Double(-
0.0536311403356718));
45 modparam[2].table_deltaH_init.put("G", new Double(-
0.787818472323185));
modparam[2].table_deltaH_init.put("T", new Double(-
0.577191417685245));

```

```

modparam[2].table_deltaH_init.put("s", new
Double(0.674274914671847));
modparam[2].table_deltaH_init.put("w", new
Double(0.737204216419899));
5
// H_mono

// H_nn
10 modparam[2].table_H_nn.put("-aa", new Double(-7.48976721304652));
modparam[2].table_H_nn.put("-at", new Double(-7.32484387195421));
modparam[2].table_H_nn.put("-ca", new Double(-7.87301403612122));
modparam[2].table_H_nn.put("-cg", new Double(-8.42170498069151));
modparam[2].table_H_nn.put("-ct", new Double(-7.74869970168917));
modparam[2].table_H_nn.put("-ga", new Double(-7.76121630019634));
15 modparam[2].table_H_nn.put("-gc", new Double(-8.50240032966233));
modparam[2].table_H_nn.put("-gg", new Double(-8.25010916492258));
modparam[2].table_H_nn.put("-gt", new Double(-7.93137471336983));
modparam[2].table_H_nn.put("-ta", new Double(-7.19824578597169));
modparam[2].table_H_nn.put("DL", new Double(-7.57500309750847));
20 modparam[2].table_H_nn.put("LD", new Double(-8.62638977709659));
modparam[2].table_H_nn.put("LL", new Double(-8.30977887038586));

// S_nn
25 modparam[2].table_S_nn.put("-aa", new Double(-21.0183983393758));
modparam[2].table_S_nn.put("-at", new Double(-20.9296841275317));
modparam[2].table_S_nn.put("-ca", new Double(-20.9035893027225));
modparam[2].table_S_nn.put("-cg", new Double(-20.6106325122583));
modparam[2].table_S_nn.put("-ct", new Double(-20.9057484730661));
modparam[2].table_S_nn.put("-ga", new Double(-20.9212322254303));
30 modparam[2].table_S_nn.put("-gc", new Double(-20.5901247977925));
modparam[2].table_S_nn.put("-gg", new Double(-20.7858543446338));
modparam[2].table_S_nn.put("-gt", new Double(-20.8689105244148));
modparam[2].table_S_nn.put("-ta", new Double(-20.9664930077991));
modparam[2].table_S_nn.put("AA", new Double(-20.9447059251337));
35 modparam[2].table_S_nn.put("AC", new Double(-21.3977869950092));
modparam[2].table_S_nn.put("AG", new Double(-19.3180598469252));
modparam[2].table_S_nn.put("AT", new Double(-21.7372825480374));
modparam[2].table_S_nn.put("Aa", new Double(-20.519746311685));
modparam[2].table_S_nn.put("Ac", new Double(-20.1102797514339));
40 modparam[2].table_S_nn.put("Ag", new Double(-19.8837896896234));
modparam[2].table_S_nn.put("At", new Double(-16.6189306374549));
modparam[2].table_S_nn.put("CA", new Double(-21.7677432772647));
modparam[2].table_S_nn.put("CC", new Double(-18.6955344509688));
modparam[2].table_S_nn.put("CG", new Double(-17.7951906152784));
45 modparam[2].table_S_nn.put("CT", new Double(-20.9908366322904));
modparam[2].table_S_nn.put("Ca", new Double(-19.5503830674945));
modparam[2].table_S_nn.put("Cc", new Double(-20.3830477876892));
modparam[2].table_S_nn.put("Cg", new Double(-18.9002469686417));
modparam[2].table_S_nn.put("Ct", new Double(-21.408268210924));

```



```

modparam[2].table_S_nn.put("GA", new Double(-20.8162756238568));
modparam[2].table_S_nn.put("GC", new Double(-20.6272454690228));
modparam[2].table_S_nn.put("GG", new Double(-18.323886570111));
modparam[2].table_S_nn.put("GT", new Double(-20.3270306928687));
5  modparam[2].table_S_nn.put("Ga", new Double(-18.8840728478426));
modparam[2].table_S_nn.put("Gc", new Double(-19.469767014314));
modparam[2].table_S_nn.put("Gg", new Double(-20.5553996930734));
modparam[2].table_S_nn.put("Gt", new Double(-19.0425869495101));
modparam[2].table_S_nn.put("TA", new Double(-22.4469367362978));
10 modparam[2].table_S_nn.put("TC", new Double(-17.8972135549923));
modparam[2].table_S_nn.put("TG", new Double(-18.5655748363087));
modparam[2].table_S_nn.put("TT", new Double(-21.8796937437075));
modparam[2].table_S_nn.put("Ta", new Double(-22.7345090588341));
modparam[2].table_S_nn.put("Tc", new Double(-20.3745455995998));
15 modparam[2].table_S_nn.put("Tg", new Double(-20.7769612524783));
modparam[2].table_S_nn.put("Tt", new Double(-21.4720373674347));
modparam[2].table_S_nn.put("aA", new Double(-20.5263443813538));
modparam[2].table_S_nn.put("aC", new Double(-19.0058736831775));
modparam[2].table_S_nn.put("aG", new Double(-20.642144617583));
20 modparam[2].table_S_nn.put("aT", new Double(-20.3395587841216));
modparam[2].table_S_nn.put("cA", new Double(-18.9418426287649));
modparam[2].table_S_nn.put("cC", new Double(-20.4556223524877));
modparam[2].table_S_nn.put("cG", new Double(-19.4702848936854));
modparam[2].table_S_nn.put("cT", new Double(-18.7713856854974));
25 modparam[2].table_S_nn.put("gA", new Double(-20.6980212771351));
modparam[2].table_S_nn.put("gC", new Double(-19.6358318414429));
modparam[2].table_S_nn.put("gG", new Double(-20.2590169785449));
modparam[2].table_S_nn.put("gT", new Double(-19.6460069771352));
modparam[2].table_S_nn.put("tA", new Double(-19.56763639202));
30 modparam[2].table_S_nn.put("tC", new Double(-19.4028441502224));
modparam[2].table_S_nn.put("tG", new Double(-21.5258339764989));
modparam[2].table_S_nn.put("tT", new Double(-21.9790781626011));

35 // =====
// Parameterfil:
/home/jgk/projects/tmpred/param/4.4.4.12.12.myhomo.ttal15

modparam[3]=new Tm_param();

40 modparam[3].table_deltaS_init=new Hashtable();
modparam[3].table_deltaH_init=new Hashtable();
modparam[3].table_H_nn=new Hashtable();
modparam[3].table_S_nn=new Hashtable();
45 modparam[3].table_H_mono=new Hashtable();

// deltaS_init
modparam[3].table_deltaS_init.put("A", new Double(-
5.14449897466861));

```

```

        modparam[3].table_deltaS_init.put("C",          new      Double(-
6.98913655256153));
        modparam[3].table_deltaS_init.put("G",          new      Double(-
5.36464180365539));
    5      modparam[3].table_deltaS_init.put("T",          new      Double(-
4.19684771688626));
        modparam[3].table_deltaS_init.put("s",          new      Double(-
0.372560237115056));
        modparam[3].table_deltaS_init.put("w",          new      Double(-
10      0.369231842748097));

        // deltaH_init
        modparam[3].table_deltaH_init.put("A",          new      Double(-
-0.308850061360838));
    15      modparam[3].table_deltaH_init.put("C",          new      Double(-
0.972721500770524));
        modparam[3].table_deltaH_init.put("G",          new      Double(-
0.856281675074907));
        modparam[3].table_deltaH_init.put("T",          new      Double(-
20      0.254870883861951));
        modparam[3].table_deltaH_init.put("s",          new      Double(0.674274914671847));
        modparam[3].table_deltaH_init.put("w",          new      Double(0.737204216419899));
    25

        // H_mono
        modparam[3].table_H_mono.put("A", new Double(-13.0850625834165));
        modparam[3].table_H_mono.put("C", new Double(-15.4415946710241));
        modparam[3].table_H_mono.put("G", new Double(-15.5290562655932));
    30      modparam[3].table_H_mono.put("T", new Double(-12.8962280546374));

        // H_nn
        modparam[3].table_H_nn.put("-aa", new Double(-7.48976721304652));
        modparam[3].table_H_nn.put("-at", new Double(-7.32484387195421));
    35      modparam[3].table_H_nn.put("-ca", new Double(-7.87301403612122));
        modparam[3].table_H_nn.put("-cg", new Double(-8.42170498069151));
        modparam[3].table_H_nn.put("-ct", new Double(-7.74869970168917));
        modparam[3].table_H_nn.put("-ga", new Double(-7.76121630019634));
        modparam[3].table_H_nn.put("-gc", new Double(-8.50240032966233));
    40      modparam[3].table_H_nn.put("-gg", new Double(-8.25010916492258));
        modparam[3].table_H_nn.put("-gt", new Double(-7.93137471336983));
        modparam[3].table_H_nn.put("-ta", new Double(-7.19824578597169));
        modparam[3].table_H_nn.put("SS", new Double(-4.94211739183454));
        modparam[3].table_H_nn.put("SW", new Double(-3.2298329995419));
    45      modparam[3].table_H_nn.put("Ss", new Double(-2.65193801394363));
        modparam[3].table_H_nn.put("Sw", new Double(-1.92736294288536));
        modparam[3].table_H_nn.put("WS", new Double(-3.74842584808002));
        modparam[3].table_H_nn.put("WW", new Double(0.250976824915542));
        modparam[3].table_H_nn.put("Ws", new Double(-2.52230884729791));

```

```

modparam[3].table_H_nn.put("Ww", new Double(-2.47894638749835));
modparam[3].table_H_nn.put("sS", new Double(-2.08573406157705));
modparam[3].table_H_nn.put("sW", new Double(-3.16193826406854));
modparam[3].table_H_nn.put("wS", new Double(-2.10629510531988));
5 modparam[3].table_H_nn.put("wW", new Double(-2.34429223089846));

// S_nn
modparam[3].table_S_nn.put("-aa", new Double(-21.0183983393758));
modparam[3].table_S_nn.put("-at", new Double(-20.9296841275317));
10 modparam[3].table_S_nn.put("-ca", new Double(-20.9035893027225));
modparam[3].table_S_nn.put("-cg", new Double(-20.6106325122583));
modparam[3].table_S_nn.put("-ct", new Double(-20.9057484730661));
modparam[3].table_S_nn.put("-ga", new Double(-20.9212322254303));
modparam[3].table_S_nn.put("-gc", new Double(-20.5901247977925));
15 modparam[3].table_S_nn.put("-gg", new Double(-20.7858543446338));
modparam[3].table_S_nn.put("-gt", new Double(-20.8689105244148));
modparam[3].table_S_nn.put("-ta", new Double(-20.9664930077991));
modparam[3].table_S_nn.put("SS", new Double(-52.7341097218716));
modparam[3].table_S_nn.put("SW", new Double(-48.0173532462563));
20 modparam[3].table_S_nn.put("Ss", new Double(-24.7053248747157));
modparam[3].table_S_nn.put("Sw", new Double(-25.2483086657963));
modparam[3].table_S_nn.put("WS", new Double(-46.7082776752559));
modparam[3].table_S_nn.put("WW", new Double(-33.9766242546579));
modparam[3].table_S_nn.put("Ws", new Double(-24.593626756038));
25 modparam[3].table_S_nn.put("Ww", new Double(-24.1779478033306));
modparam[3].table_S_nn.put("sS", new Double(-24.1344000692216));
modparam[3].table_S_nn.put("sW", new Double(-23.8911849011862));
modparam[3].table_S_nn.put("wS", new Double(-24.5041713703224));
modparam[3].table_S_nn.put("wW", new Double(-23.8041757177075));

30

// =====
// Parameterfil:
/home/jgk/projects/tmpred/param/4.4.4.12.3.myhomo.ttall6
35

modparam[4]=new Tm_param();

modparam[4].table_deltaS_init=new Hashtable();
modparam[4].table_deltaH_init=new Hashtable();
40 modparam[4].table_H_nn=new Hashtable();
modparam[4].table_S_nn=new Hashtable();
modparam[4].table_H_mono=new Hashtable();

// deltaS_init
45 modparam[4].table_deltaS_init.put("A", new Double(-
1.80759598042465));
modparam[4].table_deltaS_init.put("C", new Double(-
1.64205431364765));

```

```

    modparam[4].table_deltaS_init.put("G",      new      Double(-
1.53474825333411));
    modparam[4].table_deltaS_init.put("T",      new      Double(-
1.56722518744529));
5    modparam[4].table_deltaS_init.put("s",      new      Double(-
0.372560237115056));
    modparam[4].table_deltaS_init.put("w",      new      Double(-
0.369231842748097));

10    // deltaH_init
    modparam[4].table_deltaH_init.put("A",      new
Double(0.562011155975154));
    modparam[4].table_deltaH_init.put("C",      new
Double(0.501526805191486));
15    modparam[4].table_deltaH_init.put("G",      new      Double(-
0.223329383181387));
    modparam[4].table_deltaH_init.put("T",      new
Double(0.404061229265622));
    modparam[4].table_deltaH_init.put("s",      new
Double(0.674274914671847));
20    modparam[4].table_deltaH_init.put("w",      new
Double(0.737204216419899));

    // H_mono
25    modparam[4].table_H_mono.put("A", new Double(-17.5329265252141));
    modparam[4].table_H_mono.put("C", new Double(-18.5157490113392));
    modparam[4].table_H_mono.put("G", new Double(-18.8714990370891));
    modparam[4].table_H_mono.put("T", new Double(-17.430598429616));

30    // H_nn
    modparam[4].table_H_nn.put("-aa", new Double(-7.48976721304652));
    modparam[4].table_H_nn.put("-at", new Double(-7.32484387195421));
    modparam[4].table_H_nn.put("-ca", new Double(-7.87301403612122));
    modparam[4].table_H_nn.put("-cg", new Double(-8.42170498069151));
35    modparam[4].table_H_nn.put("-ct", new Double(-7.74869970168917));
    modparam[4].table_H_nn.put("-ga", new Double(-7.76121630019634));
    modparam[4].table_H_nn.put("-gc", new Double(-8.50240032966233));
    modparam[4].table_H_nn.put("-gg", new Double(-8.25010916492258));
    modparam[4].table_H_nn.put("-gt", new Double(-7.93137471336983));
40    modparam[4].table_H_nn.put("-ta", new Double(-7.19824578597169));
    modparam[4].table_H_nn.put("SS", new Double(-0.273099798601551));
    modparam[4].table_H_nn.put("SW", new Double(0.419286733365625));
    modparam[4].table_H_nn.put("Ss", new Double(-0.541700784403176));
    modparam[4].table_H_nn.put("Sw", new Double(0.350546387435068));
45    modparam[4].table_H_nn.put("WS", new Double(-0.296817518810437));
    modparam[4].table_H_nn.put("WW", new Double(0.175879861412384));
    modparam[4].table_H_nn.put("Ws", new Double(0.328320257160163));
    modparam[4].table_H_nn.put("Ww", new Double(-0.169293035915045));
    modparam[4].table_H_nn.put("sS", new Double(0.326542458181525));

```

```

modparam[4].table_H_nn.put("sw", new Double(-0.938829019706826));
modparam[4].table_H_nn.put("ws", new Double(0.300181620507486));
modparam[4].table_H_nn.put("ww", new Double(0.316641864778047));

5      // S_nn
modparam[4].table_S_nn.put("-aa", new Double(-21.0183983393758));
modparam[4].table_S_nn.put("-at", new Double(-20.9296841275317));
modparam[4].table_S_nn.put("-ca", new Double(-20.9035893027225));
modparam[4].table_S_nn.put("-cg", new Double(-20.6106325122583));
10    modparam[4].table_S_nn.put("-ct", new Double(-20.9057484730661));
modparam[4].table_S_nn.put("-ga", new Double(-20.9212322254303));
modparam[4].table_S_nn.put("-gc", new Double(-20.5901247977925));
modparam[4].table_S_nn.put("-gg", new Double(-20.7858543446338));
modparam[4].table_S_nn.put("-gt", new Double(-20.8689105244148));
15    modparam[4].table_S_nn.put("-ta", new Double(-20.9664930077991));
modparam[4].table_S_nn.put("DL", new Double(-22.2388993255255));
modparam[4].table_S_nn.put("LD", new Double(-23.9336954436759));
modparam[4].table_S_nn.put("LL", new Double(-48.2573001896195));

20
    // =====
    //
    // Parameterfil:
/home/jgk/projects/tmpred/param/4.4.4.48.3.myhomo.ttall3

25    modparam[5]=new Tm_param();

modparam[5].table_deltaS_init=new Hashtable();
modparam[5].table_deltaH_init=new Hashtable();
modparam[5].table_H_nn=new Hashtable();
30    modparam[5].table_S_nn=new Hashtable();
modparam[5].table_H_mono=new Hashtable();

    // deltaS_init
modparam[5].table_deltaS_init.put("A", new Double(-
35    0.888469343292043));
modparam[5].table_deltaS_init.put("C", new Double(-
1.02543607545834));
modparam[5].table_deltaS_init.put("G", new Double(-
0.541648874710394));
40    modparam[5].table_deltaS_init.put("T", new Double(-
0.65983861685686));
modparam[5].table_deltaS_init.put("s", new Double(-
0.372560237115056));
modparam[5].table_deltaS_init.put("w", new Double(-
45    0.369231842748097));

    // deltaH_init
modparam[5].table_deltaH_init.put("A", new
Double(0.615705041687039));

```

```

        modparam[5].table_deltaH_init.put("C", new
Double(0.174168418982924));
        modparam[5].table_deltaH_init.put("G", new Double(-
0.316854740516605));
5      modparam[5].table_deltaH_init.put("T", new
Double(0.468853851084089));
        modparam[5].table_deltaH_init.put("s", new
Double(0.674274914671847));
        modparam[5].table_deltaH_init.put("w", new
10 Double(0.737204216419899));

// H_mono
modparam[5].table_H_mono.put("A", new Double(-12.2836541825738));
modparam[5].table_H_mono.put("C", new Double(-12.8366937840179));
15 modparam[5].table_H_mono.put("G", new Double(-13.1042874575601));
modparam[5].table_H_mono.put("T", new Double(-12.1930059340835));

// H_nn
modparam[5].table_H_nn.put("-aa", new Double(-7.48976721304652));
modparam[5].table_H_nn.put("-at", new Double(-7.32484387195421));
20 modparam[5].table_H_nn.put("-ca", new Double(-7.87301403612122));
modparam[5].table_H_nn.put("-cg", new Double(-8.42170498069151));
modparam[5].table_H_nn.put("-ct", new Double(-7.74869970168917));
modparam[5].table_H_nn.put("-ga", new Double(-7.76121630019634));
25 modparam[5].table_H_nn.put("-gc", new Double(-8.50240032966233));
modparam[5].table_H_nn.put("-gg", new Double(-8.25010916492258));
modparam[5].table_H_nn.put("-gt", new Double(-7.93137471336983));
modparam[5].table_H_nn.put("-ta", new Double(-7.19824578597169));
modparam[5].table_H_nn.put("AA", new Double(-0.178443264646713));
30 modparam[5].table_H_nn.put("AC", new Double(-0.626677714801197));
modparam[5].table_H_nn.put("AG", new Double(-0.596829770378585));
modparam[5].table_H_nn.put("AT", new Double(-0.487815178989618));
modparam[5].table_H_nn.put("Aa", new Double(-0.771469450002954));
modparam[5].table_H_nn.put("Ac", new Double(-1.00851806983534));
35 modparam[5].table_H_nn.put("Ag", new Double(-1.42002924459608));
modparam[5].table_H_nn.put("At", new Double(-2.1602568782306));
modparam[5].table_H_nn.put("CA", new Double(-0.551742312137283));
modparam[5].table_H_nn.put("CC", new Double(-0.706523147036115));
modparam[5].table_H_nn.put("CG", new Double(-1.0219743340256));
40 modparam[5].table_H_nn.put("CT", new Double(-0.343403406607851));
modparam[5].table_H_nn.put("Ca", new Double(-1.42098576721322));
modparam[5].table_H_nn.put("Cc", new Double(-1.16189045546058));
modparam[5].table_H_nn.put("Cg", new Double(-1.50512036091287));
modparam[5].table_H_nn.put("Ct", new Double(-0.286178902087989));
45 modparam[5].table_H_nn.put("GA", new Double(-0.325781391314861));
modparam[5].table_H_nn.put("GC", new Double(0.0408233779793723));
modparam[5].table_H_nn.put("GG", new Double(-0.712259747306718));
modparam[5].table_H_nn.put("GT", new Double(-0.228069270366563));
modparam[5].table_H_nn.put("Ga", new Double(-1.53023383659402));

```

```

modparam[5].table_H_nn.put("Gc", new Double(-1.42345356134335));
modparam[5].table_H_nn.put("Gg", new Double(-1.16308519064659));
modparam[5].table_H_nn.put("Gt", new Double(-0.902484755483895));
modparam[5].table_H_nn.put("TA", new Double(-0.246725295583488));
5 modparam[5].table_H_nn.put("TC", new Double(-0.920526169197009));
modparam[5].table_H_nn.put("TG", new Double(-1.279823124502));
modparam[5].table_H_nn.put("TT", new Double(-0.317238648246969));
modparam[5].table_H_nn.put("Ta", new Double(-0.656770342200844));
modparam[5].table_H_nn.put("Tc", new Double(-0.832334138900636));
10 modparam[5].table_H_nn.put("Tg", new Double(-0.807723137936193));
modparam[5].table_H_nn.put("Tt", new Double(-0.879062052445067));
modparam[5].table_H_nn.put("aA", new Double(-0.959947989068327));
modparam[5].table_H_nn.put("aC", new Double(-1.40288154815212));
modparam[5].table_H_nn.put("aG", new Double(-0.791217583005563));
15 modparam[5].table_H_nn.put("aT", new Double(-1.02678440448276));
modparam[5].table_H_nn.put("cA", new Double(-1.72086127613672));
modparam[5].table_H_nn.put("cC", new Double(-0.706538514572194));
modparam[5].table_H_nn.put("cG", new Double(-1.27407148022265));
modparam[5].table_H_nn.put("cT", new Double(-1.67492049706038));
20 modparam[5].table_H_nn.put("gA", new Double(-1.32646498674451));
modparam[5].table_H_nn.put("gC", new Double(-1.21084101077631));
modparam[5].table_H_nn.put("gG", new Double(-1.17115932925001));
modparam[5].table_H_nn.put("gT", new Double(-1.51513536216474));
modparam[5].table_H_nn.put("tA", new Double(-0.166992658485835));
25 modparam[5].table_H_nn.put("tC", new Double(-0.957616290525477));
modparam[5].table_H_nn.put("tG", new Double(-0.724202653405165));
modparam[5].table_H_nn.put("tT", new Double(-1.08943892482195));

// S_nn
30 modparam[5].table_S_nn.put("-aa", new Double(-21.0183983393758));
modparam[5].table_S_nn.put("-at", new Double(-20.9296841275317));
modparam[5].table_S_nn.put("-ca", new Double(-20.9035893027225));
modparam[5].table_S_nn.put("-cg", new Double(-20.6106325122583));
modparam[5].table_S_nn.put("-ct", new Double(-20.9057484730661));
35 modparam[5].table_S_nn.put("-ga", new Double(-20.9212322254303));
modparam[5].table_S_nn.put("-gc", new Double(-20.5901247977925));
modparam[5].table_S_nn.put("-gg", new Double(-20.7858543446338));
modparam[5].table_S_nn.put("-gt", new Double(-20.8689105244148));
modparam[5].table_S_nn.put("-ta", new Double(-20.9664930077991));
40 modparam[5].table_S_nn.put("DL", new Double(-18.3725054920304));
modparam[5].table_S_nn.put("LD", new Double(-18.5867411854372));
modparam[5].table_S_nn.put("LL", new Double(-33.7400646113203));

45 // =====
// Parameterfil:
/home/jgk/projects/tmpred/param/4.4.4.48.3.myhomo.ttugetrel

modparam[6]=new Tm_param();

```

```

modparam[6].table_deltaS_init=new Hashtable();
modparam[6].table_deltaH_init=new Hashtable();
modparam[6].table_H_nn=new Hashtable();
5  modparam[6].table_S_nn=new Hashtable();
modparam[6].table_H_mono=new Hashtable();

// deltaS_init
modparam[6].table_deltaS_init.put("A",          new      Double(-
10  1.0899853854154));
modparam[6].table_deltaS_init.put("C",          new      Double(-
1.26650514222434));
modparam[6].table_deltaS_init.put("G",          new      Double(-
0.636096340366464));
15  modparam[6].table_deltaS_init.put("T",          new      Double(-
0.692536920626161));
modparam[6].table_deltaS_init.put("s",          new      Double(-
0.372560237115056));
modparam[6].table_deltaS_init.put("w",          new      Double(-
20  0.369231842748097));

// deltaH_init
modparam[6].table_deltaH_init.put("A",          new
Double(0.346600533782424));
25  modparam[6].table_deltaH_init.put("C",          new      Double(-
0.571391474074441));
modparam[6].table_deltaH_init.put("G",          new      Double(-
0.431395953242384));
modparam[6].table_deltaH_init.put("T",          new      Double(-
30  0.175680165547273));
modparam[6].table_deltaH_init.put("s",          new
Double(0.674274914571847));
modparam[6].table_deltaH_init.put("w",          new
Double(0.737204216419899));
35

// H_mono
modparam[6].table_H_mono.put("A", new Double(-13.0754325103332));
modparam[6].table_H_mono.put("C", new Double(-13.646666116136));
modparam[6].table_H_mono.put("G", new Double(-13.6293972843139));
40  modparam[6].table_H_mono.put("T", new Double(-12.9859631483842));

// H_nn
modparam[6].table_H_nn.put("-aa", new Double(-7.48976721304652));
modparam[6].table_H_nn.put("-at", new Double(-7.32484387195421));
45  modparam[6].table_H_nn.put("-ca", new Double(-7.87301403612122));
modparam[6].table_H_nn.put("-cg", new Double(-8.42170498069151));
modparam[6].table_H_nn.put("-ct", new Double(-7.74869970168917));
modparam[6].table_H_nn.put("-ga", new Double(-7.76121630019634));
modparam[6].table_H_nn.put("-gc", new Double(-8.50240032966233));

```



```

modparam[6].table_H_nn.put("-gg", new Double(-8.25010916492258));
modparam[6].table_H_nn.put("-gt", new Double(-7.93137471336983));
modparam[6].table_H_nn.put("-ta", new Double(-7.19824578597169));
5 modparam[6].table_H_nn.put("AA", new Double(-0.550329001961804));
modparam[6].table_H_nn.put("AC", new Double(-0.547445535909528));
modparam[6].table_H_nn.put("AG", new Double(-0.921006379530219));
modparam[6].table_H_nn.put("AT", new Double(-0.344957768635853));
modparam[6].table_H_nn.put("Aa", new Double(-0.754556907253045));
modparam[6].table_H_nn.put("Ac", new Double(-1.24531973714279));
10 modparam[6].table_H_nn.put("Ag", new Double(-1.14112776038759));
modparam[6].table_H_nn.put("At", new Double(-2.40661512922826));
modparam[6].table_H_nn.put("CA", new Double(-0.599849310599913));
modparam[6].table_H_nn.put("CC", new Double(-0.979896995845449));
modparam[6].table_H_nn.put("CG", new Double(-1.40702090497018));
15 modparam[6].table_H_nn.put("CT", new Double(-0.544368218865807));
modparam[6].table_H_nn.put("Ca", new Double(-0.84623437898934));
modparam[6].table_H_nn.put("Cc", new Double(-1.22861371811788));
modparam[6].table_H_nn.put("Cg", new Double(-1.58027912989158));
modparam[6].table_H_nn.put("Ct", new Double(-0.475151811986212));
20 modparam[6].table_H_nn.put("GA", new Double(-0.162584226406409));
modparam[6].table_H_nn.put("GC", new Double(-0.587156187857858));
modparam[6].table_H_nn.put("GG", new Double(-1.41663092804754));
modparam[6].table_H_nn.put("GT", new Double(-0.583688894933071));
modparam[6].table_H_nn.put("Ga", new Double(-1.6538514342035));
25 modparam[6].table_H_nn.put("Gc", new Double(-1.12801570402914));
modparam[6].table_H_nn.put("Gg", new Double(-0.659846271417488));
modparam[6].table_H_nn.put("Gt", new Double(-0.881015310109863));
modparam[6].table_H_nn.put("TA", new Double(-0.346920026493557));
modparam[6].table_H_nn.put("TC", new Double(-0.918176991777502));
30 modparam[6].table_H_nn.put("TG", new Double(-1.44038679704405));
modparam[6].table_H_nn.put("TT", new Double(-0.63544324592585));
modparam[6].table_H_nn.put("Ta", new Double(-0.509070031861056));
modparam[6].table_H_nn.put("Tc", new Double(-1.03584670655476));
modparam[6].table_H_nn.put("Tg", new Double(-1.40946877218105));
35 modparam[6].table_H_nn.put("Tt", new Double(-0.870845046257428));
modparam[6].table_H_nn.put("aA", new Double(-1.20602481577836));
modparam[6].table_H_nn.put("aC", new Double(-1.23960733216066));
modparam[6].table_H_nn.put("aG", new Double(-0.633900561424835));
modparam[6].table_H_nn.put("aT", new Double(-0.94390709787839));
40 modparam[6].table_H_nn.put("cA", new Double(-1.46284985286192));
modparam[6].table_H_nn.put("cC", new Double(-0.816866823421651));
modparam[6].table_H_nn.put("cG", new Double(-1.31737354533552));
modparam[6].table_H_nn.put("cT", new Double(-1.70861548243372));
modparam[6].table_H_nn.put("gA", new Double(-1.0656492536418));
45 modparam[6].table_H_nn.put("gC", new Double(-1.3976489761836));
modparam[6].table_H_nn.put("gG", new Double(-0.645745276016066));
modparam[6].table_H_nn.put("gT", new Double(-1.70832316347213));
modparam[6].table_H_nn.put("tA", new Double(-1.20474315180883));
modparam[6].table_H_nn.put("tC", new Double(-1.1530087974343));

```

```

modparam[6].table_H_nn.put("tG", new Double(-0.366342339192337));
modparam[6].table_H_nn.put("tT", new Double(-0.709379925645181));

// S_nn
5  modparam[6].table_S_nn.put("-aa", new Double(-21.0183983393758));
   modparam[6].table_S_nn.put("-at", new Double(-20.9296841275317));
   modparam[6].table_S_nn.put("-ca", new Double(-20.9035893027225));
   modparam[6].table_S_nn.put("-cg", new Double(-20.6106325122583));
   modparam[6].table_S_nn.put("-ct", new Double(-20.9057484730661));
10  modparam[6].table_S_nn.put("-ga", new Double(-20.9212322254303));
   modparam[6].table_S_nn.put("-gc", new Double(-20.5901247977925));
   modparam[6].table_S_nn.put("-gg", new Double(-20.7858543446338));
   modparam[6].table_S_nn.put("-gt", new Double(-20.8689105244148));
   modparam[6].table_S_nn.put("-ta", new Double(-20.9664930077991));
15  modparam[6].table_S_nn.put("DL", new Double(-19.0042167633551));
   modparam[6].table_S_nn.put("LD", new Double(-19.3040440770709));
   modparam[6].table_S_nn.put("LL", new Double(-37.0502230904161));

// End of model parameters
20  )

public double calcTm(String inputsequence) throws BadOligoException
{
    boolean
25  debug=Boolean.valueOf(parameterProperties.getProperty("debug","null")).
    booleanValue();
    int i;
    double[] results=new double[7];
    double retval;
30  double sum=0.0;
    int num=0;

    for(i=0; i<7; i++){
        results[i]=docalcTm(modparam[i], inputsequence);
35
        if(debug){
            System.out.println("Results[" + i + "]= " + results[i]);
        }

40
        sum+=results[i]; num++;
    }

    Arrays.sort(results);

45
    // retval=sum/num; // Mean of results
    retval=Math.round(results[3]+0.05); // Median of results (add 0.05
    to get same result as www.lna-tm.com)

    if(debug){

```

```

        System.out.println("RetVal =" + retval + " (" + results[3] + ")");
    }

    return retval; // Median of results
5    }

    private double docalcTm(Tm_param modparam, String inputsequence)
    throws BadOligoException {
        int i;
10        String mon, dimer;
        Hashtable pairs=new Hashtable();
        Hashtable endmonos=new Hashtable();
        Hashtable monos=new Hashtable();

15        double
        oligo_conc=Double.valueOf(parameterProperties.getProperty("oligo_conc",
        "null")).doubleValue();
        double
        salt_conc=Double.valueOf(parameterProperties.getProperty("salt_conc", "n
20        ull")).doubleValue();
        boolean
        debug=Boolean.valueOf(parameterProperties.getProperty("debug", "null")).
        booleanValue();

25        sequence=findReplace(inputsequence, "mC", "C");
        sequence=findReplace(sequence, "X", "C");
        if(debug){
            System.out.println("Translated seq: " + sequence);
        }

30        checkSequence(sequence);

        // Count endpairs
        String first_mon=sequence.substring(0,1);
35        String      last_mon=sequence.substring(sequence.length()-
        1,sequence.length());

        if(debug){

40        System.out.println("first_mon="+first_mon+"; "+(String)trans_ws(first_mo
        n)+"; "+(String)trans_LD(first_mon));

        System.out.println("last_mon="+last_mon+"; "+(String)trans_ws(last_mon)+
        "; "+(String)trans_LD(last_mon));
45        }

        hcount(endmonos, (String)first_mon);
        hcount(endmonos, (String)trans_ws(first_mon));
        hcount(endmonos, (String)trans_LD(first_mon));

```

```

        hcount(endmonos, (String)last_mon);
        hcount(endmonos, (String)trans_ws(last_mon));
        hcount(endmonos, (String)trans_LD(last_mon));
5
        // Count monomers
        for(i=0; i<sequence.length(); i++){
            mon=sequence.substring(i,i+1);

10            if(debug){

                System.out.println("mon="+mon+";"+(String)trans_ws(mon)+";"+(String)trans_LD(mon));
            }

15            hcount(monos, (String)mon);
            hcount(monos, (String)trans_ws(mon));
            hcount(monos, (String)trans_LD(mon));
        }

20        // Count nearest neighbors
        for(i=0; i<sequence.length()-1; i++){
            dimer=sequence.substring(i,i+2);

            if(debug){

25                System.out.println("dimer="+dimer+";"+(String)trans_ws(dimer)+";"+
                    (String)trans_LD(dimer)+";"+(String)table_ten_translate.get(dimer));
            }

            hcount(pairs, (String)dimer);
30            hcount(pairs, (String)trans_LD(dimer));
            hcount(pairs, (String)trans_ws(dimer));
            hcount(pairs, (String)table_ten_translate.get(dimer));
        }

35        // Do the calculation
        Enumeration hkeys;
        String key;

        double deltaH_sum=0, deltaS_sum=0;

40        // delta_H
        hkeys=modparam.table_deltaH_init.keys();
        while(hkeys.hasMoreElements()){
            key=(String)hkeys.nextElement();

45            if(debug){
                System.out.println("deltaH_init: "+ key + "; param=" +
                    hget(modparam.table_deltaH_init,key)+ "; val="+ hget(endmonos,key));
            }
        }

```

```

deltaH_sum+=hget(modparam.table_deltaH_init,key)*hget(endmonos,key);
    }
    hkeys=modparam.table_H_mono.keys();
5    while(hkeys.hasMoreElements()){
        key=(String)hkeys.nextElement();
        if(debug){
            System.out.println("H_mono:  "+ key+  ";  param="  +
hget(modparam.table_H_mono,key)+ ";  val="+ hget(monos,key));
10        }
        deltaH_sum+=hget(modparam.table_H_mono,key)*(hget(monos,key)-
0.5*hget(endmonos,key));
    }
    hkeys=modparam.table_H_nn.keys();
15    while(hkeys.hasMoreElements()){
        key=(String)hkeys.nextElement();
        if(debug){
            System.out.println("H_nn:  "+ key+  ";  param="  +
hget(modparam.table_H_nn,key)+ ";  val="+ hget(pairs,key));
20        }
        deltaH_sum+=hget(modparam.table_H_nn,key)*hget(pairs,key);
    }

    // delta_S
25    hkeys=modparam.table_deltaS_init.keys();
    while(hkeys.hasMoreElements()){
        key=(String)hkeys.nextElement();
        if(debug){
            System.out.println("deltaS_init:  "+ key+  ";  param="  +
30    hget(modparam.table_deltaS_init,key)+ ";  val="+ hget(endmonos,key));
        }

        deltaS_sum+=hget(modparam.table_deltaS_init,key)*hget(endmonos,key);
    }
35    hkeys=modparam.table_S_nn.keys();
    while(hkeys.hasMoreElements()){
        key=(String)hkeys.nextElement();
        if(debug){
            System.out.println("S_nn:  "+ key+  ";  param="  +
40    hget(modparam.table_S_nn,key)+ ";  val="+ hget(pairs,key));
        }
        deltaS_sum+=hget(modparam.table_S_nn,key)*hget(pairs,key);
    }

45    // Calculate Tm
    tm=new Double((double)(-273.15 + 1000
*(deltaH_sum/(deltaS_sum+0.368*(sequence.length()-
1)*Math.log(salt_conc)+1.987*Math.log(oligo_conc/4)))));

```

```
        return tm.doubleValue();  
    }  
}
```

19 MAJ 2003

Modtaget

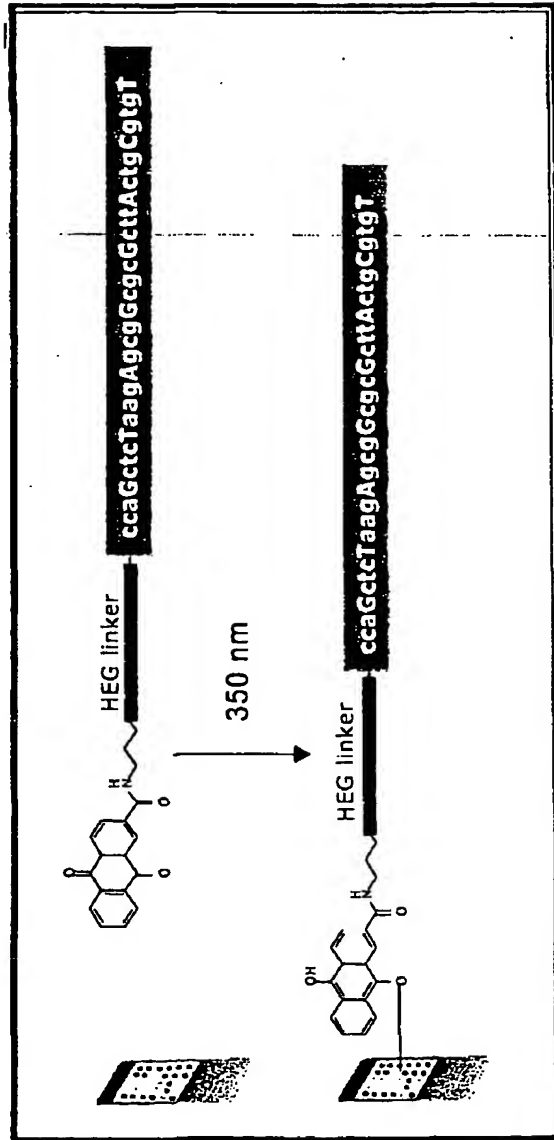


Figure 29

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

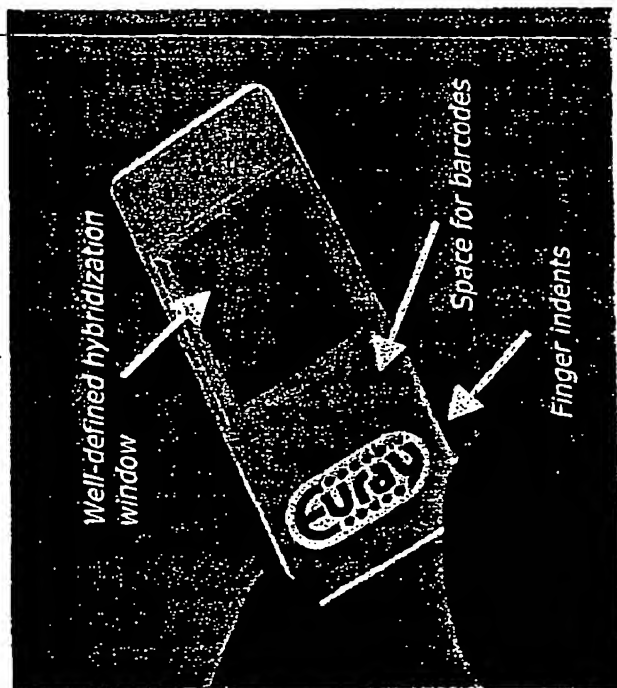


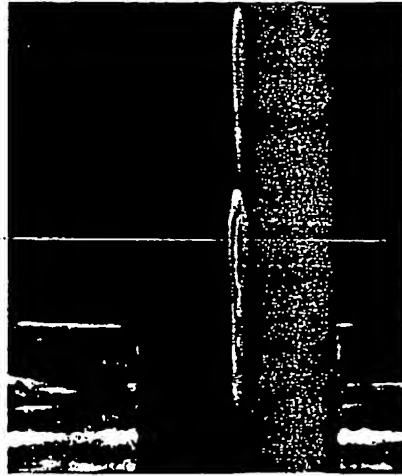
Figure 30



EURAY™ Polymer Slide



Glass Slide



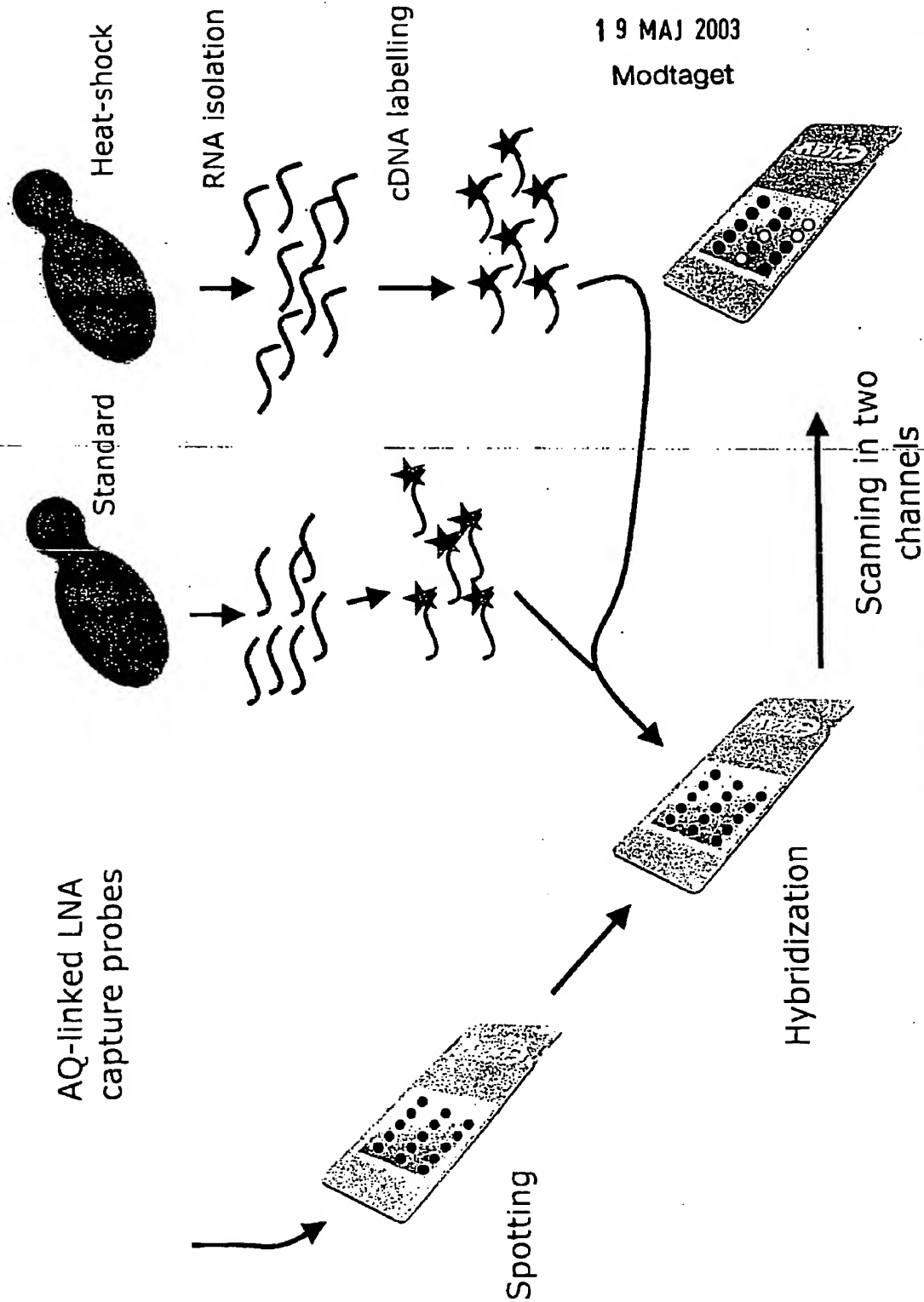
Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

Figure 3

Patent- og  
Varemærkestyrelsen  
19 MAJ 2003  
Modtaget



19 MAJ 2003

Modtaget

Gene	Encoded protein	Functional categories	mRNA copies/cell	1) Effect of heat shock	2)
ACT1	Actin	cytoskeleton	21	none	
GPD1	Glycerol-3-phosphate dehydrogenase	glycerol metabolism	7	slightly induced	
GUA1	GMP synthase	purine metabolism	8	repressed	
HAC1	Transcription factor	unfolded protein response	4	none	
HSP78	Mitochondrial	protein folding	1	induced	
RPP0	60S acidic ribosomal protein L10	protein synthesis	24	none	
SSA4	Cytosolic HSP70	ER and mitochondrial translocation	1	strongly induced	
SWI5	Transcription factor, regulates HO	cell cycle	1	slightly repressed	
THI4	Biosynthetic enzyme	thiamine biosynthesis	1	none	

Figure 33

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

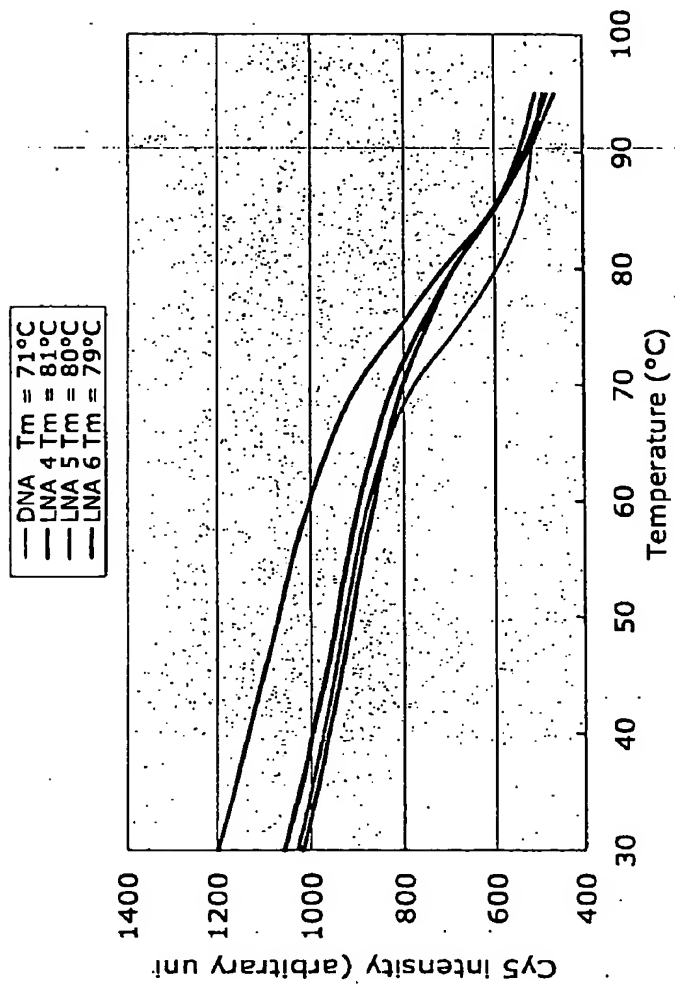


Figure 34

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

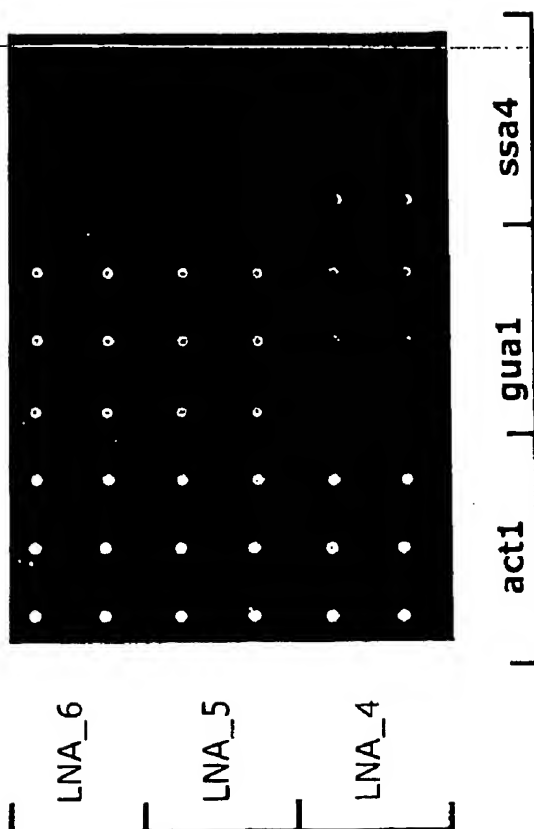


Figure 35A

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

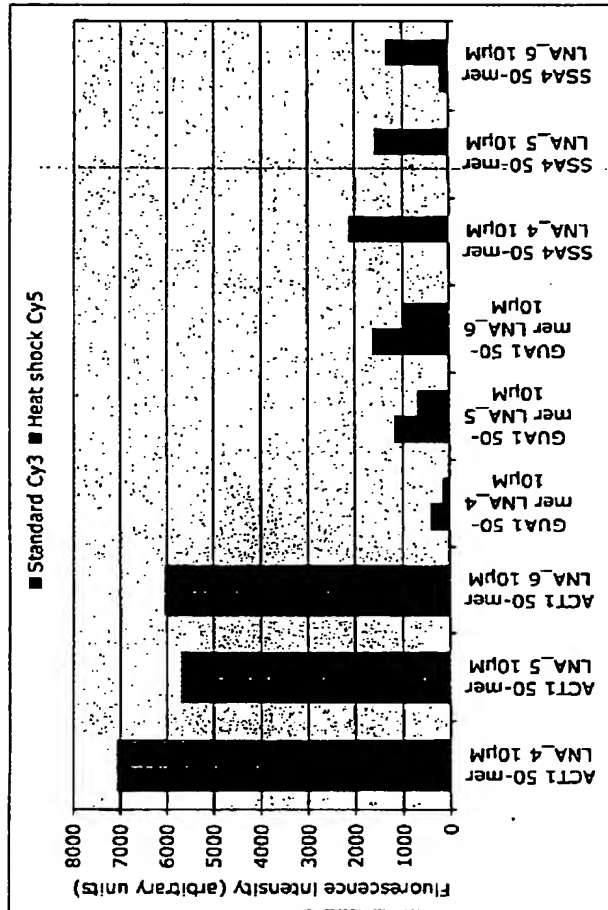
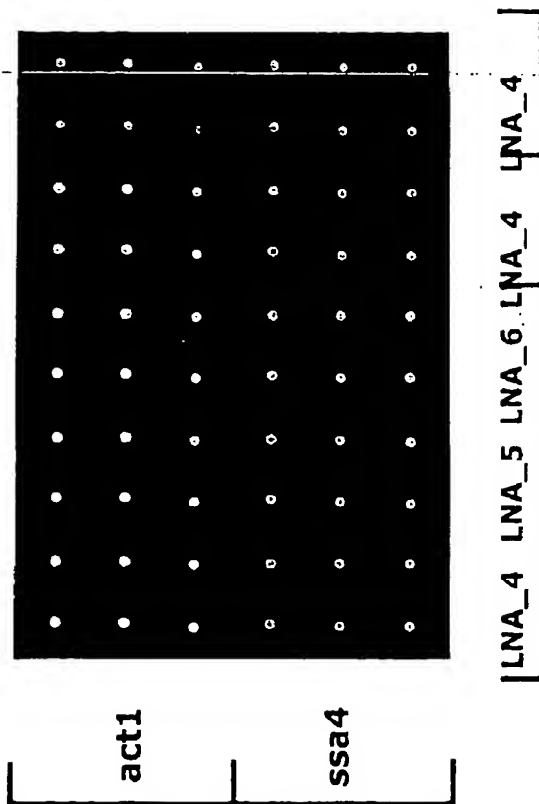


Figure 35B

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget



50-mer 40-mer 30-mer

Figure 36A

19 MAJ 2003

Modtaget

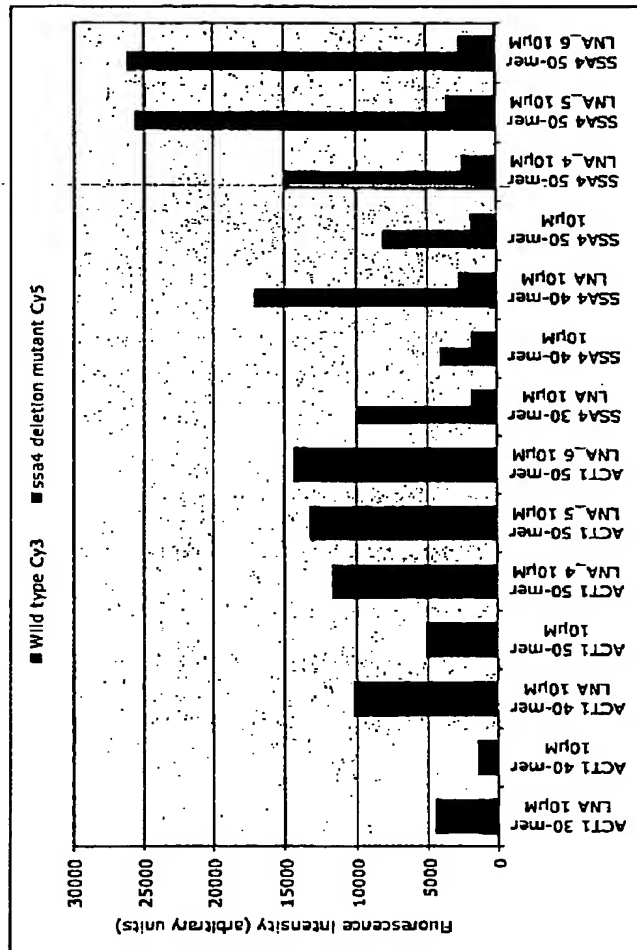
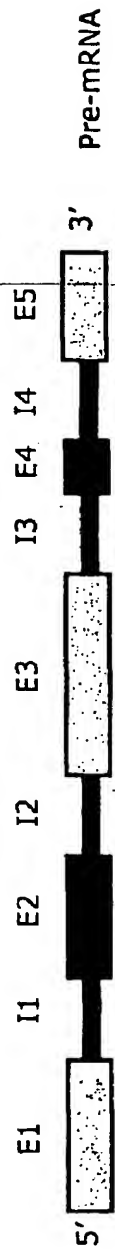


Figure 36B





Transcription



mRNA with alt.  
poly(A) sites



Exon skipping



Intron retaining



Alt. 5' splice site



Alt. 3' splice site



Figure 39.

Patent- og  
Varemærkestyrelsen

1.9 MAJ 2003

Modtaget

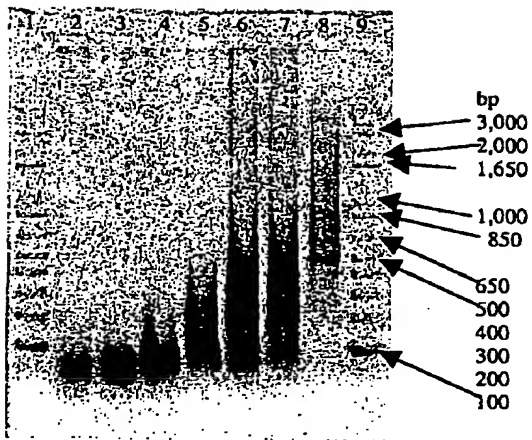


Figure 38

19 MAJ 2003

Modtaget

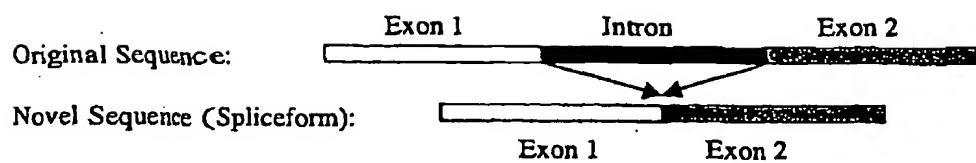


Figure 40

Patent- og  
Varemærkestyrelsen


19 MAJ 2003

Modtaget




Constant Region      Variable region

Spliceforms:

Spliceform 1: 

Spliceform 2: 

or

Spliceform 1: 

Spliceform 2: 

Figure 41

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

- Merged Probes (MP) or Junction Probes
- Unique Internal Probes (UIP)
- Shared Internal Probes (SIP)

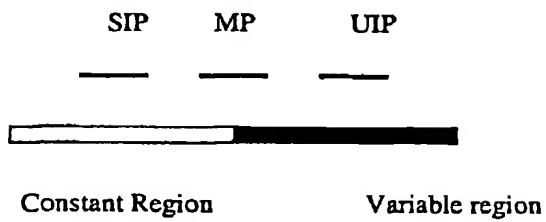


Figure 42

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

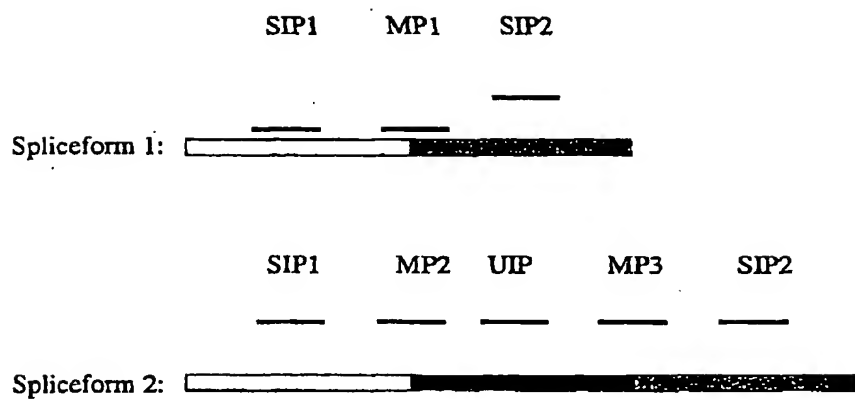


Figure 43

Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

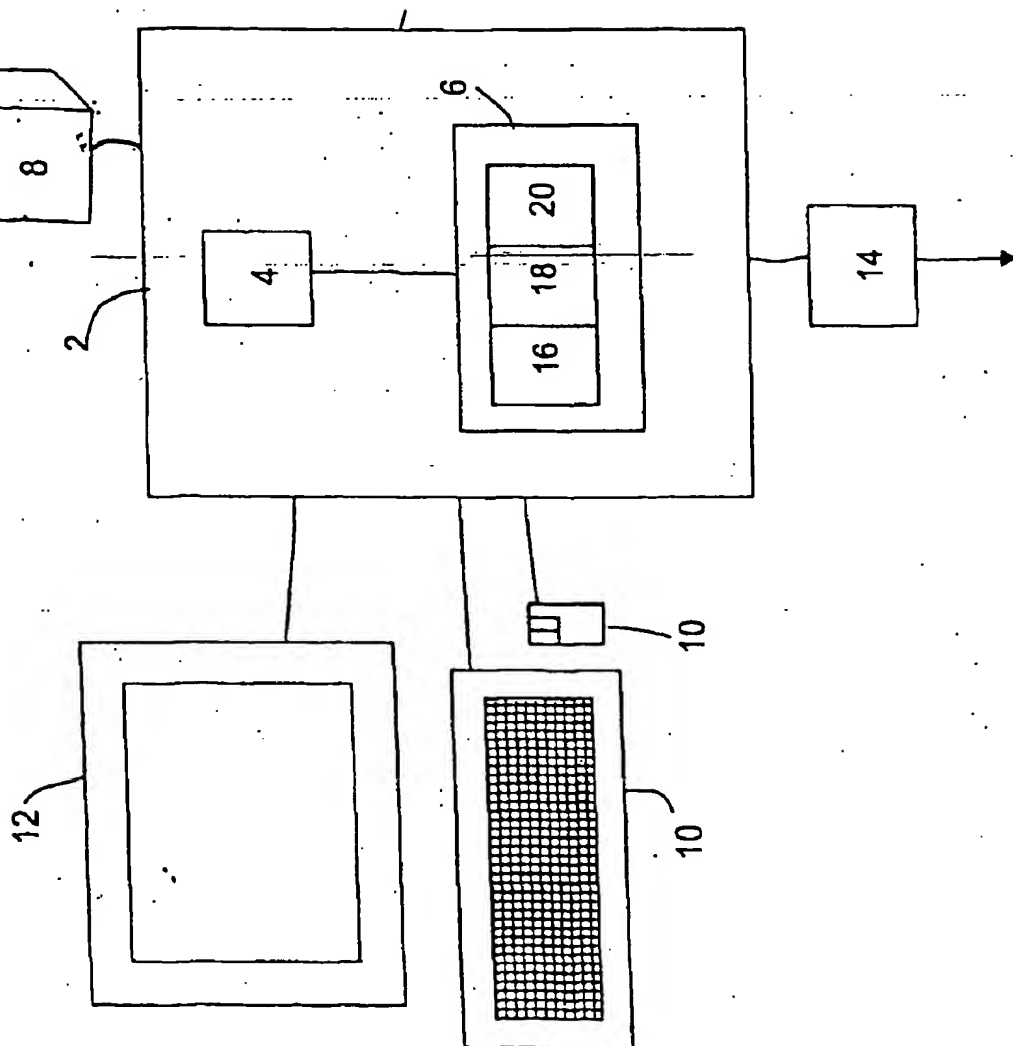


Figure 44

19 MAJ 2003

Modtaget

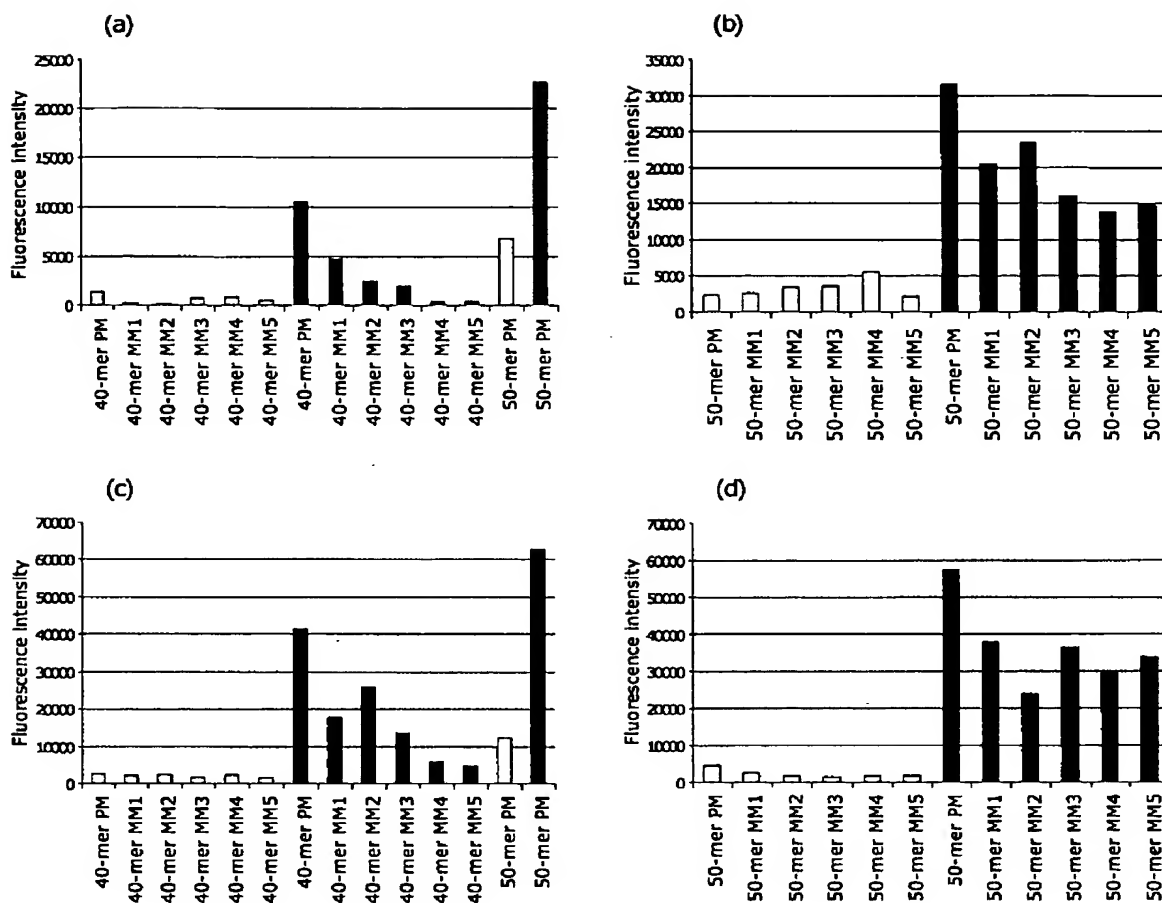


Figure 45



19 MAJ 2003

Modtaget

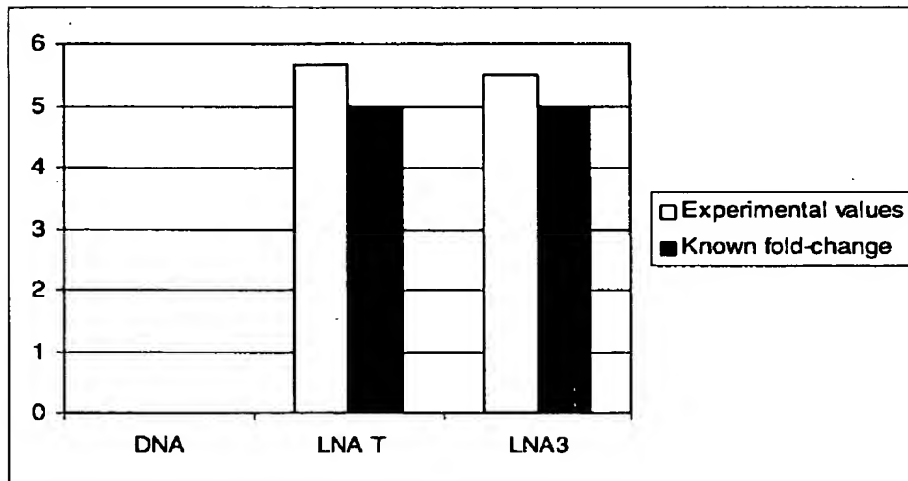


Figure 46

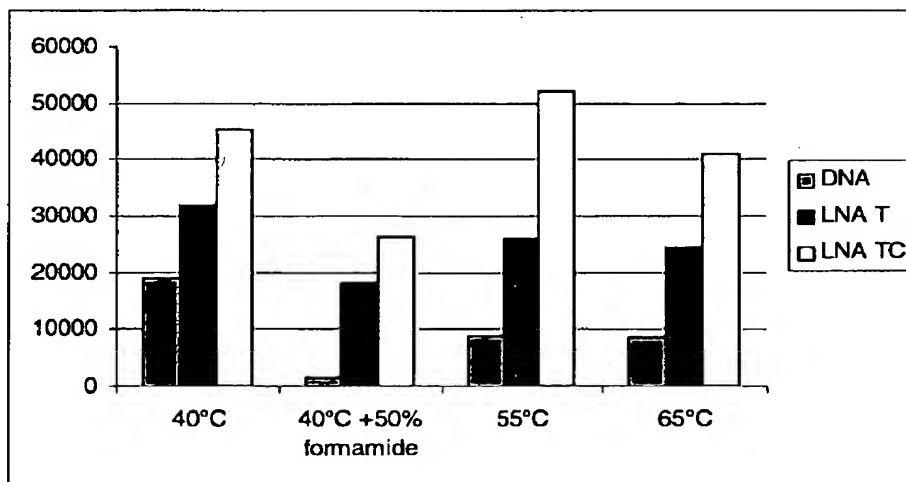


Figure 47

19 MAJ 2003

Modtaget

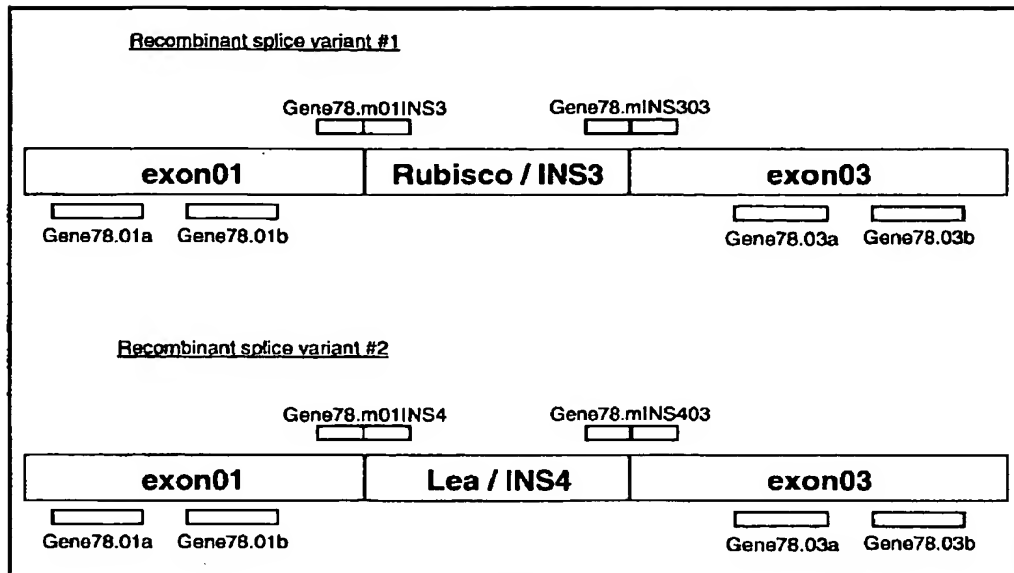


Figure 48

19 MAJ 2003

Modtaget

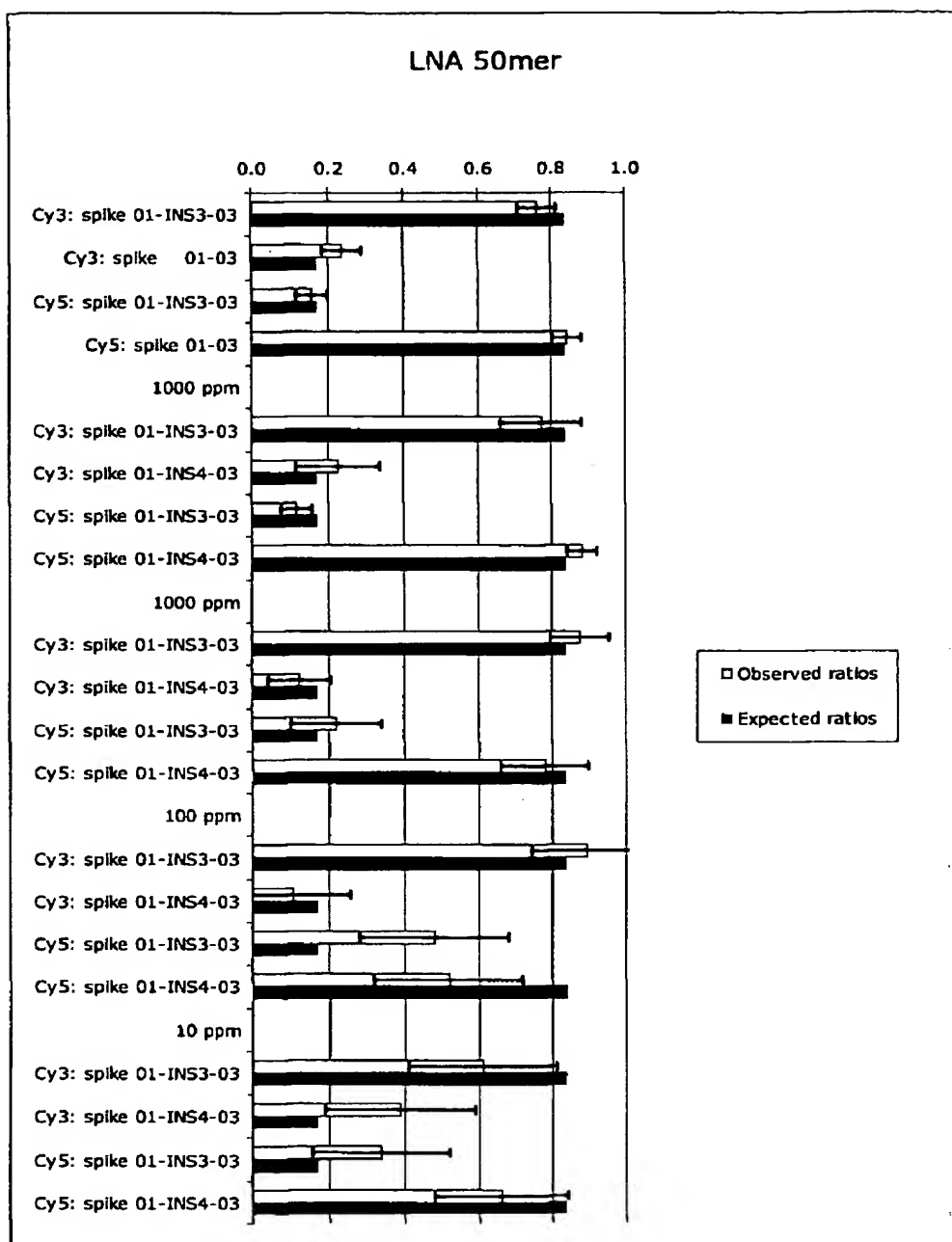


Figure 49

19 MAJ 2003

Modtaget

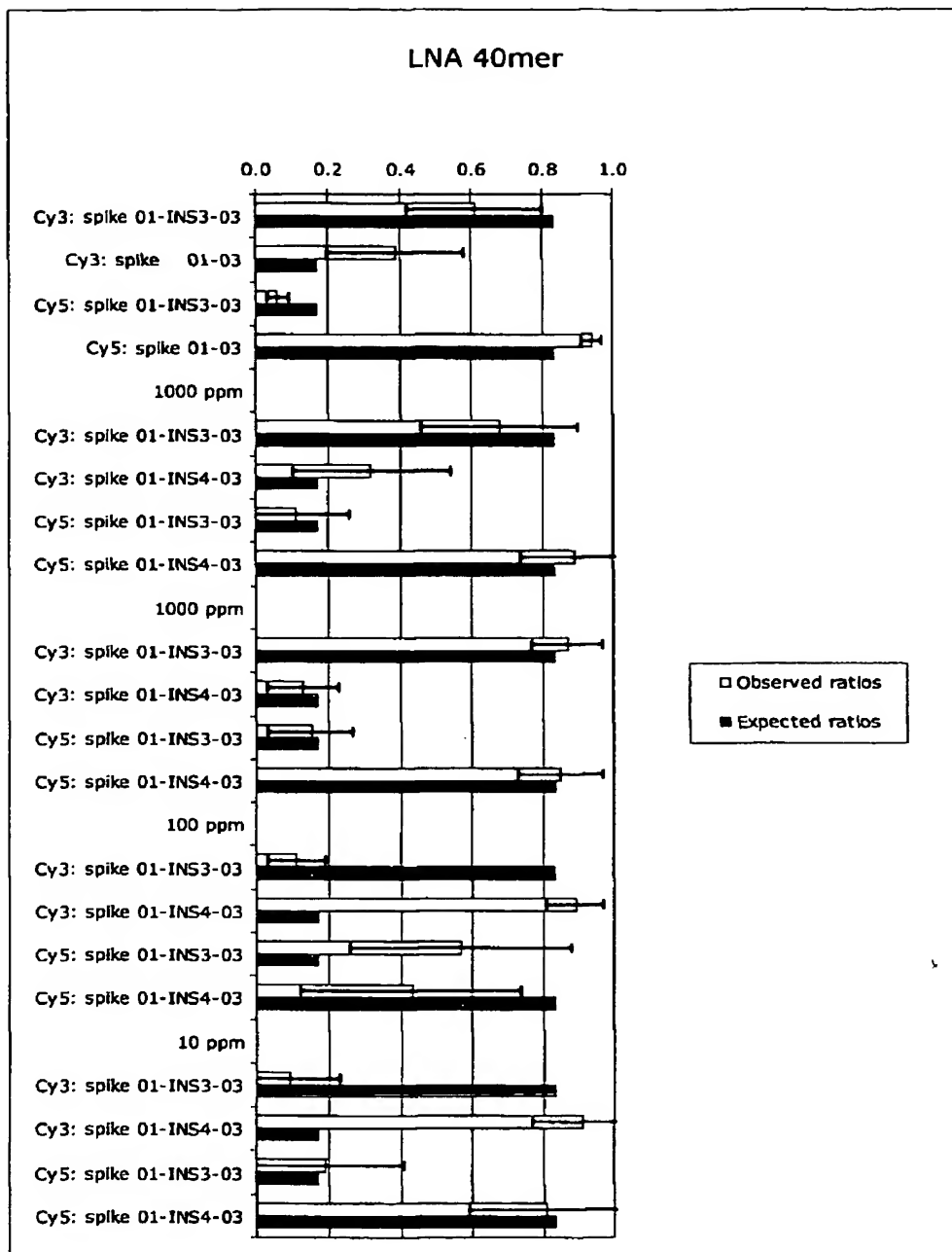
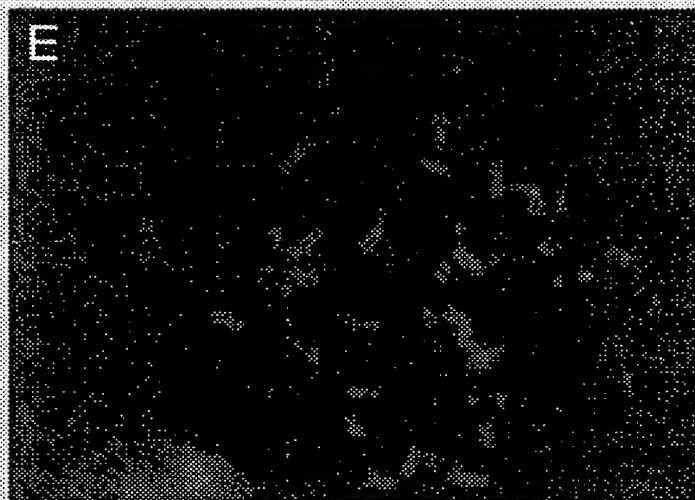
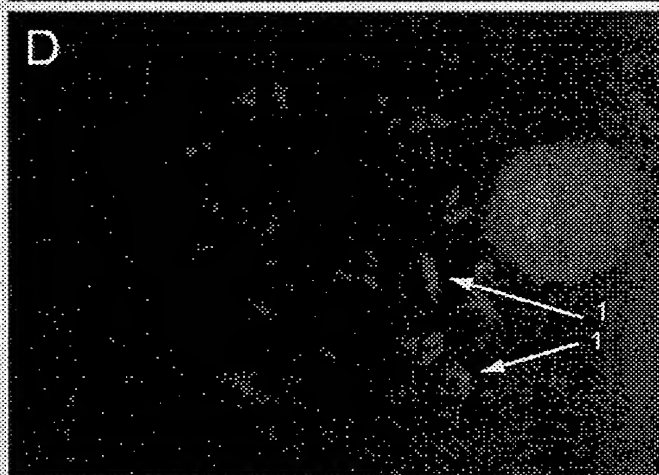
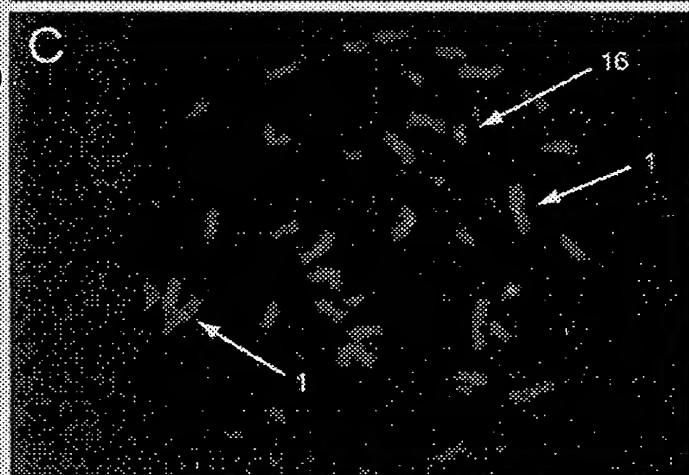
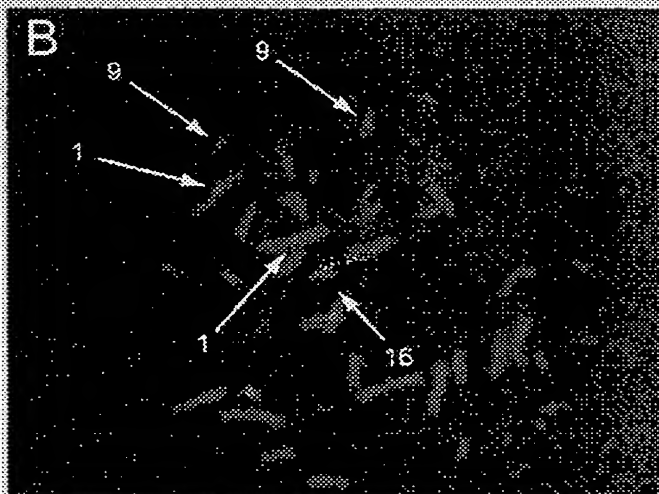
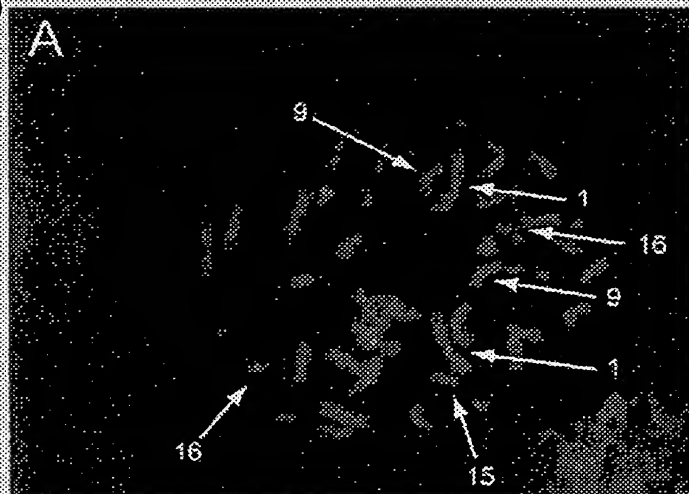


Figure 50



Patent- og  
Varemærkestyrelsen

19 MAJ 2003

Modtaget

Figure 51